TECHNISCHE UNIVERSITÄT DRESDEN

Technische Universität Dresden

Fakultät Umweltwissenschaften

Topic of the dissertation:

Model-based data worth analysis for groundwater systems with the use of surrogate models

Dissertation to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)

Submitted by:
M.Sc. Moritz Gosses
29.08.1988
Nürnberg

Reviewer	Prof. Dr. Niels Schütze, Technische Universität Dresden
Reviewer	Dr. Thomas Wöhling, Technische Universität Dresden
Reviewer	Prof. Dr. John Doherty, Flinders University

Date of defence: 11.09.2020

Erklärung des Promovenden

Die Übereinstimmung dieses Exemplars mit dem Original der Dissertation zum Thema:

"Model-based data worth analysis for groundwater systems with the use of surrogate models"

wird hiermit bestätigt.

.....
Ort, Datum

.....
Unterschrift (Vorname Name)

Abstract

The aim of this work is the improvement of model-based data worth analysis for groundwater systems with the use of surrogate models. Physically-based groundwater models are wide-spread tools used to make diverse predictions for research and management problems. They allow incorporation of system knowledge and a multitude of data. Often, this complexity is accompanied with high model run times. This is especially problematic for applications such as uncertainty analysis or data worth analysis, which necessitate many model runs. Surrogate models aim to address these challenges with run time reduction through simplification of the model, usually through techniques of the following three major categories: projection-based methods, data-driven methods and structural reduction methods. The run time reduction through the use of surrogate models is associated with impairments regarding their applicability, accuracy of system representation, predictive uncertainty quantification and integration of system knowledge. In light of these potential limitations, this thesis compares the ability of three different surrogate models in reproducing diverse model predictions and data worth estimates of a complex, real-world benchmark model. The surrogates used are a spatially and parametrically simplified physically-based model, a set of artificial neural networks (ANNs) and a projection-based "proper orthogonal decomposition" (POD) model. In the first part of this dissertation, the potentials and shortcomings of the popular POD method in regard to boundary representation are detailed and an extension of the method for accurate boundary depiction is proposed. The explicit treatment of boundary conditions is shown to eliminate reduction-induced errors at Dirichlet and Neumann boundaries (and reduce errors at Cauchy boundaries) for a small trade-off in general groundwater head accuracy. Ease of implementation and the potential for purposeful application of the extension allow modelers a target-oriented refinement of POD models. The second part of this dissertation addresses the challenge of quantifying the model simplification error of a surrogate model in light of erroneousness of their model predictions. An existing method for predictive uncertainty quantification is extended to estimate the simplification error and bias of different model predictions of all three sur-

rogate models compared to the complex benchmark model predictions. Results show that the magnitude and structure of model simplification error is highly dependent on both the type of model prediction and surrogate model. In the final part of the thesis, two of the surrogate models are compared with the complex model in their application for analysis of worth of different data types. First-order second-moment data worth analysis methods are extended to account for the non-uniqueness of calibrated model parameters in a robust method. It is then used in collaboration with the surrogate models to analyze the worth of existing, "future" and "parametric" data for varying model predictions. The comparison of changes in predictive uncertainty variance between complex model and the surrogates shows that the simplified, physically-based model is only able to identify data worth in the existing calibration data set. The POD model is a suitable surrogate for data worth analysis in regards to all differing predictions and worth of existing, "future" and "parametric" data when combined with the strengths of the proposed robust data worth analysis method.

Zusammenfassung

Das Ziel dieser Arbeit ist die Verbesserung modellbasierter Datenwertanalyse in Grundwassersystemen mittels Ersatzmodellen. Vielfältige Fragestellungen im Fachgebiet Grundwasser werden in Wissenschaft und Praxis mithilfe von physikalisch basierten Modellen beantwortet. Solche Modelle können hierbei vorhandene Kenntnisse über das System sowie eine Vielzahl gemessener Daten einbinden. Jedoch führt die zugrundeliegende Komplexität häufig zu langen Modelllaufzeiten, was insbesondere für Anwendungen mit vielen Modellläufen, wie Unsicherheits- oder Datenwertanalyse, problematisch ist. Hier greifen Ersatzmodelle an: eine Vereinfachung des Modells reduziert die zugehörigen Laufzeiten. Solche Ersatzmodelle lassen sich in drei Kategorien einteilen: projektionsbasierte Methoden, datengetriebene Methoden und strukturreduzierende Methoden. Ersatzmodelle haben jedoch auch verschiedene Nachteile: eingeschränkte Anwendbarkeit, geringere Genauigkeit der Systemabbildung, verringerte Qualität ihrer Vorhersage- und Unsicherheitsschätzung, sowie Restriktionen bei der Integrierung vorhandener Systemkenntnisse. Zur Abschätzung dieser möglichen Einschränkungen vergleicht diese Arbeit drei verschiedene Ersatzmodelle mit einem realen, komplexen Benchmarkmodell in Bezug auf ihre Fähigkeit, verschiedene Modellvorhersagen und Datenwertabschätzungen glaubhaft zu reproduzieren. Die drei Ersatzmodelle sind: ein räumlich und parametrisch vereinfachtes, physikalisch basiertes Modell, ein Verbund künstlicher neuronaler Netze und ein projektionsbasiertes "proper orthogonal decomposition" (POD) Modell. Der erste Teil der Arbeit widmet sich den Ungenauigkeiten der weitverbreiteten POD Methode bei der Abbildung von Randbedingungen und zeigt eine Erweiterung der Methode zur präziseren Darstellung der Randbedingungen in POD auf. Es wird nachgewiesen, dass diese explizite Behandlung der Randbedingungen in POD reduktionsbasierte Fehler in den Dirichlet- und Neumann-Randbedingungen eliminiert sowie Fehler in den Cauchy-Randbedingungen reduziert. Die Methode führt jedoch zu einer leicht verringerten Genauigkeit bei der allgemeinen Abbildung der Grundwasserstände im Vergleich zu herkömmlichen POD. Dennoch ist die Methode ein nützliches Werkzeug, da sie eine zielgerichtete Anpassung von POD-Ersatzmodellen erlaubt.

Der zweite Teil der Arbeit stellt sich der Aufgabe, den durch die Modellvereinfachung eingebrachten Fehler der Ersatzmodelle bezogen auf ihre Modellvorhersagen zu beurteilen. Hierzu wird eine vorhandene Methode zur Berechnung von Vorhersageunsicherheiten angepasst und erweitert, um Fehler und Bias durch Modellvereinfachung zwischen den drei Ersatzmodellen und dem Benchmarkmodell für verschiedene Modellvorhersagen abzuschätzen. Die Ergebnisse zeigen, dass Größe und Struktur des Fehlers durch Modellvereinfachung stark von der Art der Modellvorhersage und des jeweiligen Ersatzmodells abhängen. Im letzten Teil der Arbeit werden zwei der Ersatzmodelle mit dem Benchmarkmodell bezüglich ihrer Nutzbarkeit zur Datenwertanalyse verglichen. Die verwendete, lokale Methode zur Datenwertanalyse wird auf mehrere plausible Parameterfelder angewandt, um die Nichteindeutigkeit kalibrierter Modellparameter mit einzubeziehen. Mittels dieser erweiterten, robusten Methode werden vorhandene, "zukünftige" und "parametrische" Datenwerte in Bezug auf verschiedene Modellvorhersagen für die Ersatzmodelle sowie das Benchmarkmodell evaluiert. Der Vergleich der auf Vorhersageunsicherheiten bezogenen Datenwerte zwischen Benchmarkmodell und dem vereinfachten, physikalisch basierten Modell zeigt, dass dieses Ersatzmodell nur die Ergebnisse für vorhandene Daten korrekt wiedergibt. Das POD Modell dagegen ist ein geeignetes Ersatzmodell hinsichtlich seiner Fähigkeit, Datenwerte des Benchmarkmodells mittels der vorgestellten robusten Methode zu reproduzieren. Dies ist unabhängig von der gewählten Vorhersage und davon, ob vorhandene, "zukünftige" oder "parametrische" Daten betrachtet werden.

Acknowledgements

I am grateful towards many people around me, as I could not have completed this work without their help and support.

Thomas Wöhling, aka Eddy, started, fostered and shaped my academic education and career for almost ten years and through three qualifications. Thank you for your unwavering passion for science, for conveying it in infectious fashion, and for your patience, among many other great traits. I'm happy and grateful that we continue to work together.

Thanks to the Professors: Niels Schütze for welcoming me at the Institute of Hydrology despite my lack of knowledge of said topic; Wolfgang Nowak for translating complicated topics into simple words with ease; John Doherty for the extensive base of methods and knowledge, and for taking the time to review my work; and Rudolf Liedl for managing the bureaucratic tasks in a busy time.

Furthermore, I want to thank my colleagues at the Institute of Hydrology for providing open ears for questions and an open and friendly environment to work in, as well as patiently enduring many lunches while waiting for me to finish.

I want to thank Lincoln Agritech and the Marlborough District Council in New Zealand, especially Scott Wilson, for allowing me to underpin what could have been dry academic research with a real, interesting application. I'm grateful towards the "Deutsche Forschungsgemeinschaft (DFG)" for providing the financial support for this work, and the "Deutscher Akademischer Auslandsdienst (DAAD)" for allowing me to present some of it to the people most interested by it in New Zealand.

Finally, thanks to my friends and family: they provided the necessary opposite pole to work life, suffered through my ramblings about work (and life) with stoic endurance, and urged me to just sit down and finish this thesis – an unappreciated and gargantuan task at times – the urging, not the finishing.

Contents

Glossary

data worth quantification of the influence of knowledge of *observation data* on *predictive uncertainty*. Can be divided depending on the type of *observation data*:

- **existing data** available, real-world measurements.
- **"future" data** potential measurements not (yet) available.
- **"parametric" data** (existing or potential) measurements pertaining to properties implemented as *parameters* in the *model*.
- **robust [*data worth*]** analysis of *data worth* dependent on multiple (*calibrated*) *parameter* sets.

model a mathematical system that interprets *model forcings* through *parameters* to generate *model outputs*.

- **boundary conditions** mathematical specification of certain *model forcings*.
- **calibration** adjustment of *parameters* to improve fit between *observation data* and corresponding *model outputs*, i.e. *calibration data set*.
  - **[*calibration*] data set** subgroup of the *model outputs* which correspond to the *observation data* used in the *calibration*.
- **[*model*] complexity** incorporates spatial & temporal resolution, number & intricacy of *boundary conditions* and number of *parameters* of a *model*.
- **[*model*] forcings** measured or estimated system data that drives the *model*, e.g. precipitation data.
- **[*model*] outputs** *model*-generated system data, e.g. groundwater heads.
- **parameter** property of the *model* that affects its *model output* regarding to *model forcings*; pertaining to different attributes of the natural system (e.g. hydraulic conductivity).

prediction a specific *model output* of interest, usually not part of the *calibration data set*.

[*model*] realization a single instance of a *model* with its accompanying *parameter* set.

[*model*] run time time to calculate the *model outputs* for a single set of *model forcings* and *parameters*.

[*model*] structure spatial grid and dimensions and temporal resolution of a *model*.

observation data measurements of real system states pertaining to different *model outputs*; used in the *calibration* to infer *parameters*.

simplification the process of altering / exchanging a *model*, usually to reduce *model run time*.

complex model the control *model* against which the altered *surrogate models* are examined.

[*simplification*] error the discrepancy between *complex model* and *surrogate model output* due to *model simplification*.

parameter surrogacy contribution of *complex model parameters* to (often different) *surrogate model parameters*.

surrogate model result of *model simplification*: an altered / exchanged *model* of shorter *model run time* in comparison to a *complex model* (abbreviated as *surrogate*).

data-driven a method of *surrogate modeling* emulating *complex model* translation of *model forcings* to *model outputs* by learning appropriate behavior from training data sets.

projection-based a method of *surrogate modeling* projecting *complex model* equations onto a subspace generated from *model outputs*, often called snapshots.

structural simplification a method of *surrogate modeling* reducing the physical dimensions, *parameter* number and/or *boundary condition* representation of the *complex model*.

uncertainty (unknown) error of a *model*, usually of four types: model, parametric, input and measurement *uncertainty*. If not delineated otherwise, this work refers always refers to *uncertainty* under *calibration* (*postcalibration uncertainty*).

[*uncertainty*] analysis the quantitative estimation of *uncertainty* of a *model*, usually with first-order second-moment or stochastic methods.

predictive uncertainty *uncertainty* of a *model* in light of erroneous of *model predictions*.

List of Figures

List of Tables

List of Abbreviations

ANN Artificial Neural Network.

CM Complex Model.

eb-POD Proper Orthogonal Decomposition with explicit boundary treatment.

NRMSE Normalized Root-Mean Squared Error.

POD Proper Orthogonal Decomposition.

SM Surrogate Model.

$\mathbf{SM_{1,sMm}}$ Surrogate Model 1: simplified MODFLOW model.

$\mathbf{SM_{2,POD}}$ Surrogate Model 2: POD model.

$\mathbf{SM_{3,ANN}}$ Surrogate Model 3: suite of ANNs.

SVD Singular Value Decomposition.

List of Symbols

Linear algebra

$\mathbf{I}$ identity matrix.

$\mathbf{S}$ singular values.

$\mathbf{S}_1$ solution space part of the singular values $\mathbf{S}$.

$\mathbf{S}_2$ null-space part of the singular values $\mathbf{S}$.

$\mathbf{u}$ a left-singular vector.

$\mathbf{U}$ matrix of left-singular vectors.

$\mathbf{U}_1$ solution space part of the left-singular vectors $\mathbf{U}$.

$\mathbf{U}_2$ null-space part of the left-singular vectors $\mathbf{U}$.

$\mathbf{V}$ matrix of right-singular vectors.

$\mathbf{V}_1$ solution space part of the right-singular vectors $\mathbf{V}$.

$\mathbf{V}_2$ null-space part of the right-singular vectors $\mathbf{V}$.

Groundwater modeling

$\mathbf{A}$ combination of stiffness and mass matrix of a linear groundwater model.

$\mathbf{b}$ boundary conditions and external forcings.

$\mathbf{C}(\boldsymbol{\varepsilon})$ covariance matrix of the observation data set.

$\mathbf{C}(\mathbf{k})$ covariance matrix of the complex model parameters.

$\boldsymbol{\varepsilon}$ (random) measurement noise.

$\mathbf{d}$ groundwater heads at each discretized model location.

$\mathbf{h}$ real-world calibration / observation data set.

h real-world data corresponding to a model prediction.

$\mathbf{k}$ complex model parameters.

$\underline{\mathbf{k}}$ calibrated complex parameters.

$\mathbf{k}_n$ null-space contribution to the complex model parameters.

n_m number of complex model cells.

$\mathbf{o}$ complex model output.

$\mathbf{o}(\mathbf{h})$ output of the complex model calibrated against the real-world calibration data set $\mathbf{h}$.

s complex model prediction.

$s_{\mathbf{h}}$ complex model prediction calibrated against the real-world calibration data set $\mathbf{h}$.

y sensitivities of the prediction to the complex model parameters.

$\mathbf{Z}$ complex model system matrix, representing the complex model's action under calibration.

The Wairau model

f_a anisotropy ratio [–].

K_D bed conductivity of the low-land springs [m/d].

K_H hydraulic conductivity [m/d].

K_R bed conductivity of the Wairau River [m/d].

Q_{ex} river–groundwater exchange flux prediction.

Q_{sp} spring flux prediction.

$.^{sMm}$ denotation of simplified MODFLOW model ($\mathrm{SM}_{1,\mathrm{sMm}}$) parameters.

S_y specific yield [–].

S_s specific storage [m^1].

Surrogate model analysis

a slope of the regression line of the paired complex surrogate model scatter plot.

b intercept of the regression line of the paired complex surrogate model scatter plot.

$\boldsymbol{\varepsilon}_{CM}$ estimated error of the calibrated complex model prediction against reality.

$\boldsymbol{\varepsilon}_{Redu}$ contribution of complex parameter omission to the simplification error, part of vertical scatter.

$\boldsymbol{\varepsilon}_{Noise}$ effect of measurement noise on paired complex surrogate model scatter, part of horizontal scatter.

$\boldsymbol{\varepsilon}_{Null}$ null-space component of the simplification error, part of vertical scatter.

$\boldsymbol{\varepsilon}_{Surr}$ effect of parameter surrogacy on the paired complex surrogate model scatter, part of horizontal scatter.

$\mathbf{k}_o$ parameters omitted during the simplification process.

$\mathbf{k}_s$ surrogate model parameters.

$\underline{\mathbf{k}}_s$ calibrated surrogate model parameters.

$\mathbf{L}$ matrix relating surrogate model parameter set $\mathbf{k}_s$ to complex model parameter set $\mathbf{k}$ for analysis of parameter surrogacy.

$\underline{s}$ surrogate model prediction.

$\underline{s}_{\mathbf{h}}$ surrogate model prediction calibrated against the real-world calibration data set $\mathbf{h}$.

$\underline{s}_{\mathbf{o(h)}}$ prediction of the surrogate model calibrated against the (real-world calibrated) complex model output $\mathbf{o}(\mathbf{h})$.

$\hat{\underline{s}}_{\mathbf{o(h)}}$ simplification-induced-bias-corrected prediction of the surrogate model calibrated against the (real-world calibrated) complex model output $\mathbf{o}(\mathbf{h})$.

y_o sensitivities of the prediction to the omitted parameters.

y_s sensitivities of the prediction to the surrogate model parameters.

$\mathbf{Z}_o$ omitted segment of the system matrix.

$\mathbf{Z}_s$ surrogate segment of the system matrix.

Uncertainty analysis

$\cdot^{ns}$ null-space perturbed.

$\cdot^{ns\rightarrow rc}$ null-space perturbed and re-calibrated.

σ_s^2 precalibration uncertainty (as variance) of a complex model prediction s.

$\sigma_{s|\mathbf{h}}^2$ postcalibration uncertainty (as variance) of a complex model prediction s, called predictive uncertainty in this thesis.

Proper orthogonal decomposition

bc index of independent boundary conditions in the POD projection matrix $\mathbf{P}$ for eb-POD treatment.

dc index of independent Dirichlet boundary conditions in the POD projection matrix $\mathbf{P}$ for eb-POD treatment.

$\mathbf{d}_s$ complex model solution at a specific time used as part of the snapshot set $\mathbf{H}_s$.

$\mathbf{H}_s$ snapshot set comprised of n_s vectors of complex model solutions $\mathbf{d}_s$.

n_{bc} number of independent boundary condition indices.

n_r number of retained singular vectors, size of POD projection matrix.

n_{r2} size of the boundary condition-independent eb-POD projection matrix.

n_{rb} number of retained singular vectors, size of (potential) bigger POD projection matrix.

n_s number of snapshots of the complex model.

$\mathbf{N}$ matrix of collected independent boundary condition rows (bc) of the POD projection matrix $\mathbf{P}$.

$\mathbf{P}$ projection matrix for POD.

$\mathbf{Q}$ orthogonal matrix obtained from QR-factorization of the transposed independent boundary condition rows of the POD projection matrix $\mathbf{N}^T$.

$\mathbf{r}$ solution of the projected POD model.

$\boldsymbol{\Psi}$ projection matrix cleaned of boundary information for POD.

1 Introduction

Life on earth, and thus human life, depends on water as one of its major building blocks. While the seas cover about 70 % of earth's surface and hold about 97 % of the planet's water supply (Gleick, 1993), freshwater is scarcer. Freshwater is an important resource, with uses in our global economy such as drinking water supply for municipalities, process water for industries and irrigation water for agriculture, among others. Only about 1 % of freshwater is readily available as surface water in lakes and streams, and about 69 % is stored barely utilizable as ice in the caps of the polar regions and in glaciers. The remaining 30 % are groundwater in the subsurface of the earth (Gleick, 1993).

Unfortunately, groundwater is neither an omnipresent nor infinite resource. Groundwater reservoirs, called aquifers, are part of the global water cycle. Groundwater reserves are replenished through infiltration of rain, molten snow and surface waters into the ground, and naturally discharge into springs, rivers and the sea. Thus, they are influenced by natural changes to climate, soil and surface waters. Furthermore, anthropogenic activities have a huge impact on aquifers all around the world. Groundwater quantity and quality is highly impacted by human influences such as groundwater pumping for drinking water supply or agricultural irrigation, change and pollution of soils through mining and other industries, and many further interference (WWAP (UNESCO World Water Assessment Programme), 2019).

Our widespread dependency on groundwater as a freshwater source, along with the complex dependencies of groundwater quality and quantity on natural and anthropogenic influences, necessitate a sustainable management of those groundwater resources. Such management is dependent on a) gathering as much information about the natural system as possible and b) interpreting this information in regard to predictions of system behavior in past, present and future. Unfortunately, a) is hindered by the complexity and opaqueness of the subsurface, while the complexity and multitude of interactions of the natural system make b) a daunting task.

To tackle these problems, groundwater models have emerged as a major tool

for sustainable management after centuries of research of hydrogeological systems and the advent of computers. Here, a **model** describes a mathematical representation of natural processes as a basis to transfer system forcings into model outputs. This transfer from forcings to model output is governed by model structure and its parameters, which incorporate system states. The **model structure** defines the spatial and temporal resolution of the model and can range from "black-box" approaches without inherent system representation over simplified "bucket-type" structures to highly-discretized, physically-based numerical models. **Forcings** are different measured or estimated data of the natural system, like precipitation, pumping rates or river fluxes, which, implemented as **boundary conditions**, drive the behavior of the model. **Parameters** are specific attributes of the natural system that govern the transfer from forcings to model outputs. While parameters usually pertain to different attributes of the natural system, they are often hard to measure directly due to their spatial distribution, variability and location in the subsurface, among other hindrances. Instead, they are usually inferred by **calibration**, or model inversion: fitting model outputs to suitable observation data (the **calibration data set**). The **model outputs** generated by groundwater models are (spatially and temporally distributed) states such as groundwater heads, surface water fluxes or contaminant plumes. Often, measurement data exists which pertains to some of these model outputs, called **observation data**. Contrary, **predictions** are unknown model outputs of special interest for the modeler or applicants and are often the main goal of the whole modeling exercise.

Groundwater modeling is not without challenges, however. The aforementioned difficulties of data acquisition make suitable forcings and observation data scarce, expensive and uncertain. The same is true for the definition of model structure and parameterization. Theoretically, models of higher **complexity** – high parameter numbers, physically-based process representation and high spatial and temporal resolution – can be built to represent reality with a higher degree of accuracy. Practically, such increase of model complexity comes at a cost, both numerically (i.e. in the mathematical solver's ability to tackle the equation systems) and computationally (i.e. in the amount of resources and time necessary to

iteratively solve the calibration process). **Model run time** can remain a limiting factor in model construction and application, even with modern computing technology. In light of this, model **simplification** aims to reduce model run times by altering / exchanging a model to create a more simple **surrogate model** instead. Disregarding of complexity, hydrogeological models are always imperfect representations of their real-world systems. As such, predictions undertaken by these models are subject to **uncertainty** introduced through different sources, from model structure and forcings to its parameters and measurements used in the observation data. **Uncertainty analysis** to estimate the influence of these uncertainty sources, especially in light of erroneousness of model predictions (so-called **predictive uncertainty**), is thus a major field of research in groundwater modeling. Even further, **data worth** analysis methods are developed to quantify the benefit of real-world measurements on, for example, minimizing predictive uncertainty where estimates of it exist. In summary, researchers around the globe try to tackle the challenges of data availability, model uncertainty and run times, among others, to further the applicability and trustworthiness of models as tools for sustainable groundwater management.

This thesis addresses all three of the stated challenges of run times, model uncertainty and data availability, by improving model-based data worth analysis via surrogate models. Surrogate models reduce model run times, while data worth analysis evaluates measurements in light of their effect on uncertainties of model predictions. Chapter 2 introduces these topics of model uncertainty, data worth analysis and surrogate modeling in groundwater and presents the current state of the art along with open challenges and topics of scientific interest. Chapter 3 highlights how the contributions compiled in this thesis address some of those challenges. The methods applied to tackle the detailed research questions are shortly presented in chapter 4. Chapter 5 shows the results of application and discusses the suitability of surrogate models for data worth analysis for a real-world test case. Conclusions are drawn from the thesis's findings in chapter 6 and an outlook is given.

This thesis is founded on the following publications, which can be found in Appendix A:

1. Gosses, M., Nowak, W., Wöhling, T., 2018. Explicit treatment for Dirichlet, Neumann and Cauchy boundary conditions in POD-based reduction of groundwater models. Advances in Water Resources 115, 160–171. https://doi.org/10.1016/j.advwatres.2018.03.011

2. Gosses, M., Wöhling, T., 2019. Simplification error analysis for groundwater predictions with reduced order models. Advances in Water Resources 125, 4156. https://doi.org/10.1016/j.advwatres.2019.01.006

The data worth analysis with surrogates, as presented in this thesis in sections 4.5 and 5.3, is not yet published in a publication.

2 State of the art

In this chapter, I outline the research fields of surrogate modeling and uncertainty and data worth analysis in groundwater. I aim to provide an short overview of these broad research areas, identify open research questions and focus increasingly on categorizing the topics discussed in this thesis. Section 2.1 introduces the necessity for surrogate models along with mandatory terms and details the three major categories of surrogate models in groundwater. I then summarize the major open research question in the field in general and point out those addressed in this thesis. In section 2.2, I establish the concept of uncertainty in groundwater modeling along with its sources and summarize the types of uncertainty quantification methods common. Furthermore, I introduce the concept of data worth analysis and examine its use in groundwater research. Again, I close by recapitulating the main open research questions and contextualize the issues addressed in this thesis accordingly.

2.1 Surrogate modeling for groundwater systems

Presenting the topic of surrogate modeling in groundwater research, I first provide an overview of the necessity for and advantages and disadvantages of surrogate models (2.1.1). I then provide explanation of the basic principles behind the three major categories of surrogate modeling in groundwater research: projection-based methods (2.1.2), data-driven methods (2.1.3) and structural simplification methods (2.1.4). For each category, I discuss advantages and disadvantages along with applications and open topics in current research. Finally, I summarize the open research questions in groundwater surrogate modeling and integrate the relevant topics addressed by my work (2.1.5).

2.1.1 Introduction

As introduced earlier, the need for surrogate models in groundwater usually stems from one drawback of complex, "realistic" groundwater models: model run time. Generally, the advantages of highly parameterized, physically-based groundwater

models are manifold: spatially distributed outputs, use of all available system knowledge in regard to process and boundary representation, ability to predict model outputs for future or management scenarios, and more. Unfortunately, among other possible disadvantages, this can come with prohibitively high model run times. A major research area trying to mitigate this problem is the use of surrogate models in groundwater research.

As the major criteria for the analysis of surrogate models in groundwater, Asher et al. (2015) declare in their review paper that surrogates should "significantly increase computational efficiency", highlighting the importance of this objective. The efficiency is most often measured by comparison of run times between complex and surrogate model (e.g. Burrows and Doherty, 2015; Cardoso et al., 2009; Ushijima and Yeh, 2015), while some studies also incorporate potential complex model runs necessary for the surrogate model construction (e.g. Pasetto et al., 2013; Vermeulen et al., 2004a). Note that in this thesis, I use the term **complex model** as the control model against which accompanying surrogate models are benchmarked. The run time reduction through surrogate models can be used to accelerate different modeling tasks such as forecasting (e.g. Adamowski and Chan, 2011; Daliakopoulos et al., 2005), inverse modeling (e.g. Boyce and Yeh, 2014; Burrows and Doherty, 2016; Winton et al., 2011), stochastic simulation (e.g. Pasetto et al., 2013), uncertainty analysis (e.g. Burrows and Doherty, 2016; Doherty and Christensen, 2011) and optimization of management strategies (e.g. McPhee and Yeh, 2008; Mulligan and Ahlfeld, 2016) or experimental design (e.g. Ushijima and Yeh, 2015). Besides alleviating this major drawback of complex models, high model run times, surrogate models can also possibly reduce numerical instability (Doherty and Christensen, 2011), aid in analysis of (in-)sensitive parameterization (Watson et al., 2013) or be used for the determination of adequate model structures (Matott et al., 2009).

Proper evaluation of the performance of a surrogate model necessitates another measure besides the run time reduction compared to the complex model, which, as detailed above, shows the benefit associated with using the surrogate model: a comparison of the performance between surrogate model and complex

model in regards to their representation of the system of interest. This can be viewed as a **simplification error**: a measure of discrepancy between outputs of complex model and surrogate model due to the simplification. Estimation of this simplification error reveals the drawback of the surrogate model. Ideally, the simplification error should be minimal, but it is hard to estimate and highly dependent on its respective definition. Furthermore, other possible disadvantages are associated with the use of surrogate models, too. Oftentimes, the implementation might not be straightforward (e.g. Boyce et al., 2015), the models are dependent on certain state spaces or parameter sets highly limiting their value (e.g. McPhee and Yeh, 2008), or are altogether limited to linear applications (e.g. Vermeulen et al., 2004a). In addition, certain surrogates might be prone to overfitting (Razavi et al., 2012a) or struggle with capturing heterogeneity (Najm, 2009).

Many researchers tackle these contrasts of run time reduction and erroneousness of surrogate models in groundwater research with a wide variety of model simplifications or stand-ins. To facilitate the literature review, I distinguish three major categories that most surrogate models can be assigned to, following Asher et al. (2015): projection-based methods, data-driven methods and structural reduction methods. The overarching methodology of each category is described in the following, different applications in groundwater research are listed and advantages and disadvantages of each category of surrogate models are discussed. Table 1 summarizes some of the recent literature studies in groundwater research for all three categories.

2.1.2 Projection-based methods

The first surrogate model category consists of the projection-based methods. **Projection-based methods** of surrogate modeling take the governing equations of the complex model and project them onto a subspace of smaller dimension for numerical solution. The original model solution is approximated by a linear combination of basis vectors defining the projection and the calculated subspace solutions. The category includes methods such as proper orthogonal decomposition (POD) (e.g. Boyce and Yeh, 2014; Siade et al., 2010; Vermeulen et al., 2004a),

Karhunen-Loève expansion (e.g. Laloy et al., 2013), Krylov subspace methods (e.g. Woodbury et al., 1990) or Fourier model reduction (e.g. Willcox and Megretski, 2005), among others. They mainly differ in their construction of the basis vectors used for the subspace projection and approximation of the complex model solution.

The most common projection-based method in groundwater modeling is POD. The method builds the basis vectors from a collection of complex model solutions incorporating the variability of the complex model. Those solutions, the so-called snapshots, are decomposed into orthonormal vectors through solution of an eigenvalue problem (e.g. Pasetto et al., 2011; Siade et al., 2012; Vermeulen et al., 2004b) or singular value decomposition (SVD) (e.g. Boyce and Yeh, 2014; Stanko et al., 2016). They then serve as basis vectors for the subspace projection, and can further be reduced (in number) through principal component analysis (PCA) due to their inherent ranking (e.g. Vermeulen et al., 2004b; Siade et al., 2010).

Groundwater research in the past decade has focused on different aspects of the POD method. Siade et al. (2010) determined methods for snapshot selection to generate snapshot sets of adequate complex model representation efficiently. Boyce and Yeh (2014) and Ushijima and Yeh (2015) applied POD surrogate models to the solution of inverse problems. Boyce et al. (2015) and Stanko et al. (2016) worked on extending the method for non-linear applications, while Gosses et al. (2018) focused on the treatment of boundary conditions in POD.

Projection-based methods implicitly keep the system knowledge embedded in the complex model's mathematical formulation, boundary representation and parameterization through extraction of the basis vectors from complex model solutions and projection of the complex model formulation. They provide a significant reduction in model run times (e.g. Pasetto et al., 2013). Furthermore, simplification errors through projection-based surrogate modeling are shown to be small for linear applications (e.g. Vermeulen et al., 2004b). Unfortunately, non-linearities (both in regard to boundaries and parameters) are common in groundwater models and highly limit the accuracy and/or computational efficiency of projection-based methods (Vermeulen et al., 2006). Appropriate snapshot selection is another dif-

ficulty of those methods – while the snapshot set has to incorporate all variability of the complex model that the surrogate should reproduce, snapshot creation requires (long) complex model runs, thus limiting the overall efficiency of the surrogate. In addition, the implementation of projection-based methods is usually quite complex, as the governing equations of the complex model have to be altered.

2.1.3 Data-driven methods

The second category contains data-driven methods, which are often also called response surface surrogates in literature. Razavi et al. (2012b) defines **data-driven methods** of surrogate modeling as "statistical or empirical data-driven models emulating the high-fidelity model responses". Generally speaking, data-driven methods do not simulate the physical processes of the considered system, but do emulate the relationship between forcings and required outputs by "learning" the appropriate translation through training against a measurement data set. They can also be used in conjunction with a complex model, where they simulate model outputs of interest (predictions) based on knowledge of model parameters and model forcings, thus functioning as "metamodels of original models" (Razavi et al., 2012b).

Different data-driven methods have been used widely in hydrological applications, such as artificial neural networks (e.g. Taormina et al., 2012), Bayesian networks (e.g. Fienen et al., 2013) or polynomial chaos methods (e.g. Marzouk and Najm, 2009). They have been successfully used for the prediction of groundwater heads in literature (e.g. Daliakopoulos et al., 2005; Yoon et al., 2011). In their capability as metamodels, they are often applied to aid calibration and optimization problems (e.g. Khu and Werner, 2003; Razavi et al., 2012a).

A major advantage of many data-driven methods is the huge computational time saving often associated with them (cf. Gosses and Wöhling, 2019; Khu and Werner, 2003). On the one hand, due to their "black-box" nature, exact knowledge of the underlying physical processes is not necessary for their application. On the other hand, their omission of process representation might exclude available data

and system knowledge. As data-driven methods accumulate appropriate system knowledge solely from the training data without the restraints of typical physical groundwater models, this can lead to overfitting (Razavi et al., 2012a). Large problem spaces (in the number of parameters for metamodel approaches) can be problematic, as well (Razavi et al., 2012b).

2.1.4 Structural simplification methods

The last surrogate model category consists of structural simplification methods. Calling models using **structural simplification methods** "multifidelity based surrogates", Asher et al. (2015) define them as "constructed from the complex model by reducing numerical resolution, increasing tolerances, or removing processes". I also incorporate model simplification through reduction of parameters into this category.

Structural simplification methods are mainly used in conjuncture with the complex model counterparts on which they are built. Doherty and Christensen (2011) and Watson et al. (2013) proposed a paired simple complex model approach to analyze uncertainties associated with the surrogate model, which was simplified both spatially and parametrically. Von Gunten et al. (2014) used surrogate models based on grid-coarsening to accelerate the calibration process of a large-scale hydrological catchment model. Multiscale models (e.g. Sun, 2008) can be included here, as the attempt to link different models at different scales leads to a simplification in comparison to building a complex model from the fine scale for the whole domain.

Structural simplification is often the easiest method to implement, as changes to model grid and parameters can usually be made directly in the chosen modeling software. Furthermore, the computational cost savings can be easily estimated beforehand, as single model run time scales roughly proportionally with the number of grid cells and (linear) calibration iterations with the number of parameters. Parameter reduction can be done following a sensitivity analysis, ensuring the heterogeneity in hydrogeological representation necessary for predictions of interest is preserved. Coupling of surrogate and complex model can reveal strengths and

weaknesses of the surrogate model, thus leading to a more robust comprehension of possible applications (Doherty and Christensen, 2011; Watson et al., 2013). Multiscale methods allow different physical representations of processes at different scales, and can potentially be highly parallelized (Sun, 2008). In contrast, structural simplification directly counteracts the advantages of physically-based models to begin with. A model grid too coarse can lead to numerical artifacts in the solution. Parameter reduction can severely inhibit the representation of hydrogeological heterogeneity, negatively influencing model predictions depending on it. The mapping of parameters between surrogate and complex model is not straightforward (cf. Watson et al., 2013), possibly limiting applications for inverse modeling or uncertainty analysis.

2.1.5 Open research questions

As presented above and aggregated in Table 1, the research questions in the field of surrogate modeling in (ground-) water science are diverse, both across and within the three categories. Many researchers tackle underlying conceptual issues of the respective methods, like appropriate snapshot selection or non-linear extensions for POD. Others apply surrogate models for quicker model calibration or uncertainty analysis, or use them for forecasting. The applications range from simple, synthetic 1D test cases for methodological research to complex 3D groundwater models simulating real-world systems on basis of existing input and measurement data.

Nonetheless, many questions remain unclear in the field of surrogate modeling in groundwater. I address two different topics in this thesis: the major one is the application of different surrogate models from all three major categories for a single real-world case as an attempt to compare the surrogate models for different applications. This incorporates extensive analysis of the model simplification error in regard to predictive performance to ensure an extensive comparison. Furthermore, I attend to the erroneousness of the boundary treatment of the POD method and present an extension to alleviate this problem.

Table 1: Surrogate models in recent hydrological literature

Category	Study	Method	Application	Research question
Projection-based methods	Siade et al. (2010)	POD	3D real-world groundwater model	Snapshot selection
	Laloy et al. (2013)	Karhunen-Loève expansion	3D real-world groundwater model	Non-linear calibration and uncertainty analysis
	Boyce and Yeh (2014)	POD	1D & 2D (semi-) synthetic groundwater models	Calibration
	Ushijima and Yeh (2015)	POD	1D & 2D synthetic groundwater models	Experimental design
	Stanko et al. (2016)	POD	1D & 2D synthetic groundwater models	Extension for non-linear problems
Data-driven methods	Yoon et al. (2011)	ANNs, Support Vector machines	Real-world groundwater levels	Forecasting
	Taormina et al. (2012)	ANNs	Real-world groundwater levels	Forecasting
	Fienen et al. (2013)	Bayesian networks	Real-world groundwater levels	Forecasting
Structural simplification methods	Doherty and Christensen (2011)	Spatial & parametrical simplification	3D synthetic groundwater model	Uncertainty analysis
	Watson et al. (2013)	Spatial & parametrical simplification, lumped parameter model	3D synthetic groundwater model	Uncertainty analysis, parameter surrogacy
	von Gunten et al. (2014)	Spatial simplification	3D real-world hydrological model	Calibration

2.2 Uncertainty and data worth analysis

In this section, I combine the topics of uncertainty and data worth analysis, as I focus on examining worth of data through the lens of affecting (predictive) uncertainties. I start of with a short introduction into uncertainty analysis in hydrosciences (2.2.1) followed by a classification of the different sources of uncertainty encountered in groundwater modeling (2.2.2). Then, I briefly present the two major schools of though regarding uncertainty analysis in hydrosciences along with their respective advantages and disadvantages (2.2.3). I follow with a quick overview of data worth analysis in regards to uncertainty analysis (2.2.4) and end with a summary of open research questions and contextualize the topics addressed in this thesis (2.2.5).

2.2.1 Introduction

Groundwater models are, regardless of their complexity, always imperfect representations of reality. This is due to the impossibility to fully capture and recreate the real system. Thus, groundwater models entail uncertainties of various sources (as discussed below). Naturally, analysis of these uncertainties associated with groundwater models and their predictions has been a highly investigated topic for several decades now (e.g. Beven and Binley, 1992; Dausman et al., 2010; Kunstmann et al., 2002; Schöniger et al., 2015). Researchers tackle varied questions regarding uncertainty in groundwater modeling: analyzing uncertainty dependent on its source (e.g. Beven and Binley, 1992; Kavetski et al., 2006), through a geological lense (e.g. Refsgaard et al., 2012), regarding practical design or management questions (e.g. Cirpka et al., 2004; Hall and Solomatine, 2008) or even more philosophical examinations (e.g. Pappenberger and Beven, 2006). In the following, I focus on guiding the review towards the categorization of the uncertainty and data worth analysis methods used in this thesis.

2.2.2 Sources of uncertainty in groundwater modeling

In consideration of uncertainty in groundwater modeling it is vital to identify the source of the uncertainty. This allows a more thorough examination of its magnitude and form, its effects on model outputs and if and how it could be reduced. In groundwater modeling, it is common to distinguish between four major sources of uncertainty (e.g. Refsgaard and Storm, 1996):

1. Model (or structural) uncertainty, which is introduced through the overall model setup and structure that is chosen. This is complicated to address, as it usually requires the setup of multiple (different) models. The uncertainty associated with boundary conditions is often lumped into this model uncertainty term. For examples of research about model uncertainty, see Doherty and Christensen (2011), Schöniger et al. (2015) and von Gunten et al. (2014), among others. I do not directly discuss model uncertainty in this thesis, though the simplification error analysis is methodologically somewhat similar to a model uncertainty analysis. Instead of estimating model uncertainty in comparison to an unknown real world, it estimates the simplification error in comparison to a known complex benchmark model, though.

2. Parametric uncertainty, which incorporates that model parameters can not be determined precisely. Usually, model parameters represent certain physical parameters of the real-world. Unfortunately, these real physical parameters can usually not be measured directly, nor are they identical to the model parameters. Thus, model parameters are often determined through inverse calibration against measurements of different system states. For complex numerical groundwater models, this process of inverse model calibration is ill-posed (Beven and Binley, 1992), meaning that there are several possible parameter combinations that would all sufficiently calibrate the model in regard to the chosen observation data set. This means that even in light of perfect fit in a calibration exercise, a single parameter set is only part of many possible combinations spaning the parameter set – the parametric

uncertainty. While this parametric uncertainty can be estimated by using stochastic simulation for joint model calibration and uncertainty analysis (see Keating et al. (2010), for example), the huge amount of model runs necessary for such methods is usually not feasible. I incorporate parametric uncertainty in the data worth estimation process presented in this thesis, as is shown later.

3. Input uncertainty, which stems from model forcing data such as rainfall data or river gaugings (e.g. Kavetski et al., 2006). Such forcing data is usually generated from real-world measurements (which are subject to measurement errors), often spatially interpolated (which results in errors), or even taken from other model's outputs (which are erroneous). All of these factors lead to uncertainty in the model input through its forcings. Where future scenario calculation is of interest, the input uncertainty is magnified by our limited knowledge of future system behavior. In this thesis, I do not address input uncertainty.

4. Measurement uncertainty, which is connected to the observed system states used as calibration targets in the model. Such measurements are obviously subject to (unknown) measurement errors. Furthermore, many system states of interest, such as groundwater heads, are continuous in space and time in reality, but measured at certain, distinct points in space and time. Thus, even potentially perfect measurements portray very specific circumstances. Measurement error is often represented by certain (semi-random) statistical distributions directly (cf. Hill and Tiedeman, 2006). Research such as Cooley and Christensen (2006) show that such assumptions are not always correct, though. I treat measurement uncertainty only indirectly, i.e. through representation in weighing of the observation data in calibration data sets, in this thesis.

To conclude, there are four different sources of uncertainty in groundwater models. In the following work, I do not analyze model or input uncertainties. Measurement uncertainty is incorporated through certain statistical assumptions, while

the methods presented below aim to estimate the effects of parametric uncertainty on predictive uncertainty and data worth estimations by accounting for parameter non-uniqueness.

2.2.3 Types of uncertainty analysis

Uncertainty analysis in hydro(geo-)logical research is usually following one of two modes of thought: first-order second-moment uncertainty analysis (e.g. Cirpka et al., 2004; Doherty, 2016; James and Oldenburg, 1997; Kunstmann et al., 2002) and stochastic uncertainty analysis (e.g. Beven and Binley, 1992; James and Oldenburg, 1997; Kavetski et al., 2006; Keating et al., 2010).

First-order second-moment uncertainty analysis tries to adequately assess the erroneous associated by the uncertainty types of interest and propagate these errors through the model onto the prediction. The easiest case would be the statistical definition of measurement error (as mean and variance of a Gaussian distribution, for example) and linear propagation of such measurement errors from the calibration data set to the model prediction (e.g. Cirpka et al., 2004; Kunstmann et al., 2002). The major advantage of these methods is the low computational demand associated with their application. Unfortunately, propagation of errors through the model is based on assumptions of propagation models and linearity which are often inaccurate, leading to the introduction of (unknown) error into the uncertainty estimate.

Stochastic uncertainty analysis computes uncertainties through continued sampling of the possible solution space. Usually, uncertainties are defined as distributions, which are sampled repeatedly by methods such as Monte-Carlo sampling (Beven and Binley, 1992; James and Oldenburg, 1997). These methods are not hampered by any non-linearity, and only require certain assumptions about the prior distribution of model parameters and input and measurement errors. Unfortunately, sampling distributions has a high computational demand, often prohibitively so, despite modern sampling methods severely increasing efficiency (e.g. Kavetski et al., 2006; Keating et al., 2010).

Some research is focused on comparing uncertainty estimates using both meth-

ods. (Keating et al., 2010) found that new methods like null-space Monte-Carlo showed promise efficiently sampling the parametric uncertainty without the use of computationally expensive Monte-Carlo methods.

2.2.4 Data worth analysis

Many researchers in hydrological sciences have taken on the question of how different data affect uncertainties in modeling. Most often, worth of data is estimated in regards to reducing/increasing the uncertainty associated with model predictions (e.g. Beven and Binley, 1992; Cirpka et al., 2004; Dausman et al., 2010). Throughout this thesis, I utilize this same viewpoint of data worth in regards to predictive uncertainty. Note that I do not discuss the topic of network design, which also associates costs of potential measurements (e.g. Erickson et al., 2002; Feyen and Gorelick, 2005) in this thesis.

Worth of data can be estimated for different types of data. In this thesis, I differentiate between three different types, loosely following Fienen et al. (2010): existing data, "future" data and "parametric" data. Dausman et al. (2010); Fienen et al. (2010) estimated the worth of existing data points, thus computing the increase in predictive uncertainty through the potential exclusion of existing measurements. Frequent is the estimation of "future" data worth (e.g. Fienen et al., 2010; Wöhling et al., 2016). The applied methods, one of which is detailed later in this thesis, allow an estimation of the potential reduction of predictive uncertainty through the knowledge of currently unknown, yet-to-be gathered measurements of system states. This can be used to identify areas for and types of measuring that strongly decrease uncertainty of key predictions, used to advise measurers in the field to take target-oriented measurements. "Parametric" data refers to potential knowledge of certain model parameters, and its effect on reducing predictive uncertainty (e.g. Feyen and Gorelick, 2005; Moore and Doherty, 2006; Fienen et al., 2010). This can be used to identify parameters highly sensitive to predictions of interest.

In analyzing worth of data for predictive uncertainty estimation, researchers have applied many different methods for the underlying predictive uncertainty

estimation, both first-order second-moment methods (e.g. Cirpka et al., 2004; Dausman et al., 2010; Fienen et al., 2010) and stochastic methods (e.g. Feyen and Gorelick, 2005; Fu and Gómez-Hernández, 2009). Thus, such data worth analysis comes with the same implications and restrictions as the utilized uncertainty quantification methods: stochastic analysis is often unfeasible due to the high number of model runs required, while first-order second-moment based analysis is subject to (unknown) error. A number of authors have tried to mitigate those problems by using methods that employ targeted sampling of the parameter space via null-space Monte-Carlo, for example (Siade et al., 2017), or by reducing run times of global search algorithms through the use of surrogate models (Schultz et al., 2006).

2.2.5 Open research questions

As presented above, research in uncertainty analysis in hydrosciences deals with different uncertainty sources via different methods. Both input and measurement uncertainty regard external data sources that are often not the main focus of groundwater modelers and are thus often researched by their respective fields (such as meteorologists, for example). Model uncertainty is very complex to analyze, usually involving several different models. Parametric uncertainty can potentially be estimated very well via stochastic methods, but these are very time-consuming. Thus, researchers focus on approximating such results via less cost-intensive approaches. In this thesis, I present a method to combine existing first-order second-moment uncertainty and data worth analysis methods with null-space Monte-Carlo techniques to account for parametric uncertainty without the time requirements of full stochastic methods. I apply this robust data worth analysis method to benchmark and surrogate models to test the ability of the surrogate models to reproduce data worth estimates of the complex model with this framework.

3 Objectives and contributions

This thesis addresses the applicability of surrogate models for predictive uncertainty and data worth analysis in coupled surface water–groundwater systems. The assessment of uncertainty of model predictions is a vital part of groundwater modeling in research and management. This can be extended by analysis of existing, "future" and "parametric" data in regards to their benefit in reducing the uncertainty of model predictions of interest. Application of these methods is often hindered by prohibitively long model run times, though.

Surrogate models are a possible solution to alleviate for long model run times, thus allowing implementation of more robust uncertainty and data worth assessment for a multitude of model predictions. Unfortunately, model simplification is always accompanied by loss of information. The location, magnitude and extension of this information loss is highly dependent on the system in question, the prediction of interest and the type of surrogate model used. Furthermore, the popular surrogate modeling method POD lacks in proper treatment of boundary conditions, which is of high importance for the application in coupled surface water–groundwater systems. Therefore, this thesis engages in the following research questions:

1. Can the implicit error in treating boundary conditions in POD be improved?
2. Which types of surrogate models are suitable to reproduce different complex model predictions with a small model simplification error?
3. Can surrogate models be used to reproduce worth of data estimated by a complex model via a robust uncertainty analysis method?

3.1 Explicit boundary treatment in POD

POD is a popular surrogate modeling technique in groundwater research over the last decade. While many of its potentials and shortcomings have been discussed in literature, the handling of model boundary conditions is of major interest in the light of coupled surface water–groundwater modeling. Therefore, the first part of

this thesis focuses on the quantification of boundary accuracy in POD surrogate modeling (see sections 4.2 and 5.1). I analyze the boundary errors of the existing POD method for all three main types of groundwater model boundaries. I find the errors to be significant and correlated even for simple, small-scale synthetic test models. To address these problems, I successfully adopt a simple method of explicit treatment of Dirichlet boundaries in POD modeling to the field of groundwater models, eliminating errors at those boundaries. Furthermore, I extend the method to both Neumann- and Cauchy-type boundaries, which can thus be satisfied completely error-free (Neumann) or made more accurate (Cauchy). The resulting trade-off of overall groundwater head representation of the POD model vs. accuracy at the boundaries can be easily quantified as shown in Gosses et al. (2018). Therefore, I provide modelers with a tool to allow informed decision-making on the boundary-specific accuracy of POD-based model simplification.

3.2 Analysis of model simplification error

The second part of this thesis is devoted to the quantification of model simplification error for different surrogate models and model predictions (see sections 4.3, 4.4 and 5.2). In Gosses and Wöhling (2019), we adopt and extend an existing method of complex–surrogate model coupling to compare the predictive performance of three distinct surrogate models to a complex, real-world benchmark model (CM). The surrogate models I use for this analysis are a spatially and parametrically simplified physically-distributed groundwater model ($SM_{1,sMm}$), a projection-based "proper orthogonal decomposition" ($SM_{2,POD}$) model and a set of artificial neural networks ($SM_{3,ANN}$). Complex model and surrogate models are coupled by creation of a large set of semi-random complex model realizations, which outputs are used as "observations" for subsequent model calibration of all three surrogate models. I create scatter plots depicting predictive bias and error of the surrogate models using sets of model predictions computed by both the random complex model realizations and the calibrated surrogate model realizations. Testing of the resulting confidence intervals shows the $SM_{3,ANN}$ models to be formidable reproducers of predictions contained in the calibration data set,

while the $SM_{1,sMm}$ model can be used to generate predictions of acceptable bias and error in most cases. The $SM_{2,POD}$ model also shows mixed results, performing better where predictions are similar in type and location to the calibration data set. This thorough comparison of the predictive quality of three different surrogate models identifies strengths and weaknesses for each surrogate modeling type for different predictions under the light of application to groundwater–surface water models.

3.3 Robust data worth analysis using surrogate models

As detailed in sections 4.5 and 5.3, the final part of the thesis extends the analysis of predictive performance of surrogate models to the question of data worth. A robust method of data worth analysis is constructed through the combination of first-order second-moment methods with efficient sampling of the calibrated parameter space, utilizing the potential for many model runs due to short surrogate model run times. I use the concept of null-space parameter perturbation to generate a multitude of complex model realizations that all satisfy the observation data set and then undertake localized first-order second-moment data worth analysis for existing, "future" and "parametric" data for each realization. The increase (through subtraction of existing data) or reduction (through addition of potential "future" measurements or "parametric" knowledge) of predictive uncertainty is computed for the complex model and two of the surrogates: the $SM_{1,sMm}$ model and the $SM_{2,POD}$ model. I show that the $SM_{1,sMm}$ model is only able to reproduce the worth of existing data while failing to correctly predict locations of worth for both "future" and "parametric" data for different predictions. The $SM_{2,POD}$ model is generally able to identify locations of worth independent of the analysis of existing, "future" and "parametric" data for all predictions. The high variance associated with the individual $SM_{2,POD}$ model results suggests the importance of the proposed robust scheme to combine first-order second-moment data worth methods with use of multiple parameterizations to account for parametric uncertainty.

3.4 Expected impact

I see diverse use of the different innovations of this work in the area of surrogate groundwater modeling. The explicit boundary treatment is an easy-to-implement method to improve accuracy of POD application in groundwater disregarding the specific research question or management application. The simplification error analysis presented allows the examination of appropriateness of chosen surrogates for specific targets, which is extended to the possibility of surrogate data worth analysis through the methods provided here.

4 Methods

In this chapter, I present the models under investigation and the methods of implicit boundary treatment in POD, simplification error analysis and robust data worth analysis with surrogate models applied in this thesis.

In section 4.1, I present the Wairau aquifer (4.1.1), which was used as the area of investigation in this work, and the complex MODFLOW model (4.1.2) used as a starting point and benchmark for the surrogate models. The setup of the three different surrogate models stemming from the three major surrogate modeling categories is described in this section as well: a simplified MODFLOW model (4.1.3), a POD model (4.1.4) and a suite of ANNs (4.1.5).

In section 4.2, I focus on the enhancement of POD, a popular surrogate modeling method in groundwater, for better boundary condition treatment. I shortly summarize the mathematical background of POD surrogates in groundwater modeling and explain its shortcomings regarding boundary representation (4.2.1). Then, I present the theory of explicit boundary treatment in POD (4.2.2) and discuss its application to the three types of boundary conditions in groundwater modeling (4.2.3). I end the chapter with some general comments on the cost associated with this method in comparison to basic POD (4.2.4).

As investigation of the error associated with surrogate model use is highly important for their responsible application, I present a method for analyzing the simplification error. This description of the simplification error analysis is split in two sections: section 4.3 deals with the theoretical background of the simplification error analysis. I present the mathematical definition of a linear model along the concepts of solution space and null-space (4.3.1) and give a mathematical definition of a surrogate model (4.3.2). I summarize the concept of parameter surrogacy and establish a process to relate complex and surrogate model parameters (4.3.3). Finally, I show the mathematical derivation of the simplification error (4.3.4).

The second part of the simplification error analysis, the method of graphical estimation via scatter plots, is presented in section 4.4. First, I present the methodology for the creation of the scatter plots (4.4.1) used in analysis of the

simplification error. Then, I explain its general features (4.4.2), the representation of the different error terms (4.4.3) defined in the theoretical background and the concept of prediction pairs (4.4.4). At the end, I summarize the main concepts of the scatter plot analysis of model simplification error (4.4.5).

The last section of this chapter, section 4.5, contains the methods for uncertainty and data worth analysis implemented in this thesis. I start by definition of first-order second-moment pre- and post-calibration uncertainty of a model prediction (4.5.1), with the latter being used as predictive uncertainty throughout this work. I then present the concept of worth of data in context of this predictive uncertainty (4.5.2), regarding existing, "future" and "parametric" data. Furthermore, I show the process of efficient generation of calibrated parameter sets via null-space perturbation of a calibrated parameter set (4.5.3). Finally, I connect this process with the concepts of data worth analysis presented earlier in the section to outline the robust data worth analysis applied in this thesis (4.5.4).

4.1 The Wairau Plain aquifer, the complex model and its surrogates

This section delineates the real-world case study used throughout the work in this thesis, along with its modeling applications, in shortened form. Subsection 4.1.1 gives a short overview of the research area. In subsection 4.1.2, I describe the complex MODFLOW model which serves as a benchmark for the simplification error analysis and the data worth results of the surrogate models. All three different surrogate models are characterized shortly in this section: the simplified MODFLOW model, $SM_{1,sMm}$, in subsection 4.1.3, the POD model, $SM_{2,POD}$, in subsection 4.1.4, and the suite of artificial neural networks, $SM_{3,ANN}$, in subsection 4.1.5. A full-length description of the Wairau aquifer and the complex coupled surface water–groundwater model can be found in Wöhling et al. (2018), while two of the three different surrogate models are presented in detail by Gosses and Wöhling (2019).

4.1.1 The Wairau Plain aquifer

The theoretical work in this thesis was applied to an existing research area: the Wairau Plain aquifer in New Zealand. Located in the northern part of the South Island near the city of Blenheim, the Wairau Plain aquifer is a shallow gravel aquifer under extensive utilization as both a drinking supply and for agricultural irrigation. The aquifer is situated under an approximately 27 km long floodplain of the Wairau River, is between 20–35 m thick and consists of an upper and lower gravel aquifer, both highly permeable, which are intermitted by a clay-rich gravel layer of lower permeability. The Wairau River flows along the northern model boundary. Leakage of river water in the western and north-western model domain through the riverbed is the main contributor to aquifer recharge in the model. A confining, wedge-shaped layer called Dillon's Point lies on top of the aquifer in the eastern part of the model, which pushes the groundwater towards the surface through several springs and back into the north-eastern part of the Wairau River.

Four river gauging stations provide flow data for the Wairau River, and discharge of the main low-land spring, called Spring Creek, is measured with a flow-recorder. Distributed over the Wairau Plain are several groundwater head observation wells, run partly continuous, partly intermittedly. A virtual climate station network operated by NIWA (Tait et al., 2006) provides precipitation and potential evapotranspiration data. Pumping from the manifold of wells situated on the Wairau Plain is estimated, along with the groundwater recharge, by a soil–water balance model (Wöhling et al., 2018).

4.1.2 Complex MODFLOW model of the Wairau Plain aquifer (CM)

The research of simplification errors of surrogate models and their capabilities in estimating data worth undertaken in this thesis necessitates a complex benchmark model as a point of comparison for the surrogate models. This complex model of the Wairau Plain aquifer that I use as the benchmark is a transient coupled surface water–groundwater model realized in MODFLOW-NWT (Niswonger et al., 2011) using the ModelMuse (Winston, 2009) GUI. It consists of 6,360 active model cells, spanning the approx. 85 km^2 model area with a 200 $\times$ 200 m horizontal

grid and three vertical layers representing the major geological units, as shown in Figure 1a. Data from 01.07.2013 to 06.11.2016 was used to spin up, calibrate and validate the model, resulting in a total of 1,225 daily time steps. The eastern, downstream model boundary was set as a constant groundwater flux boundary simulating the groundwater flux into the sea. The northern boundary, set by the Wairau River, and the southern boundary, a no-flow boundary set perpendicular to estimations of the groundwater flow field, meet in the western, upstream end of the model domain. The Wairau River was implemented via the stream-flow routing package (Niswonger and Prudic, 2005), the springs in the eastern model area were simulated as drains, and the groundwater recharge and abstraction estimates of the soil–water balance model were used as model forcings over the domain.

90 pilot point parameters each were used to discretize both hydraulic conductivity K_H and specific yield S_y of the three geological units for a total of 180 parameter values. The interpolation of these values followed an exponential variogram derived from pumping tests, and regularization techniques were used to ensure consistency with this variogram for the pilot point values. Furthermore, for each of the three geological units, a single specific storage parameter S_s and anisotropy ratio f_a was defined. For the Dillon's Point aquitard, a single parameter was used for each of these four geological properties. Finally, bed conductivities of the Wairau River (K_R) and for the different springs (K_D) were defined as parameters, for a total of 207 model parameters of the full MODFLOW model, as summarized in Table 2.

Five continuous groundwater head measurements (labeled 1–5) were used for parameter calibration, along with measured discharge at the Spring Creek (SC) spring. Differential Wairau River flow gaugings and estimates of mean discharge for the other springs were used as "soft" targets during the model calibration. Model predictions range from groundwater heads at other measurement points (labeled 6–10) to piece-wise river–groundwater exchange fluxes and spring discharges. These are presented in more detail in Gosses and Wöhling (2019).

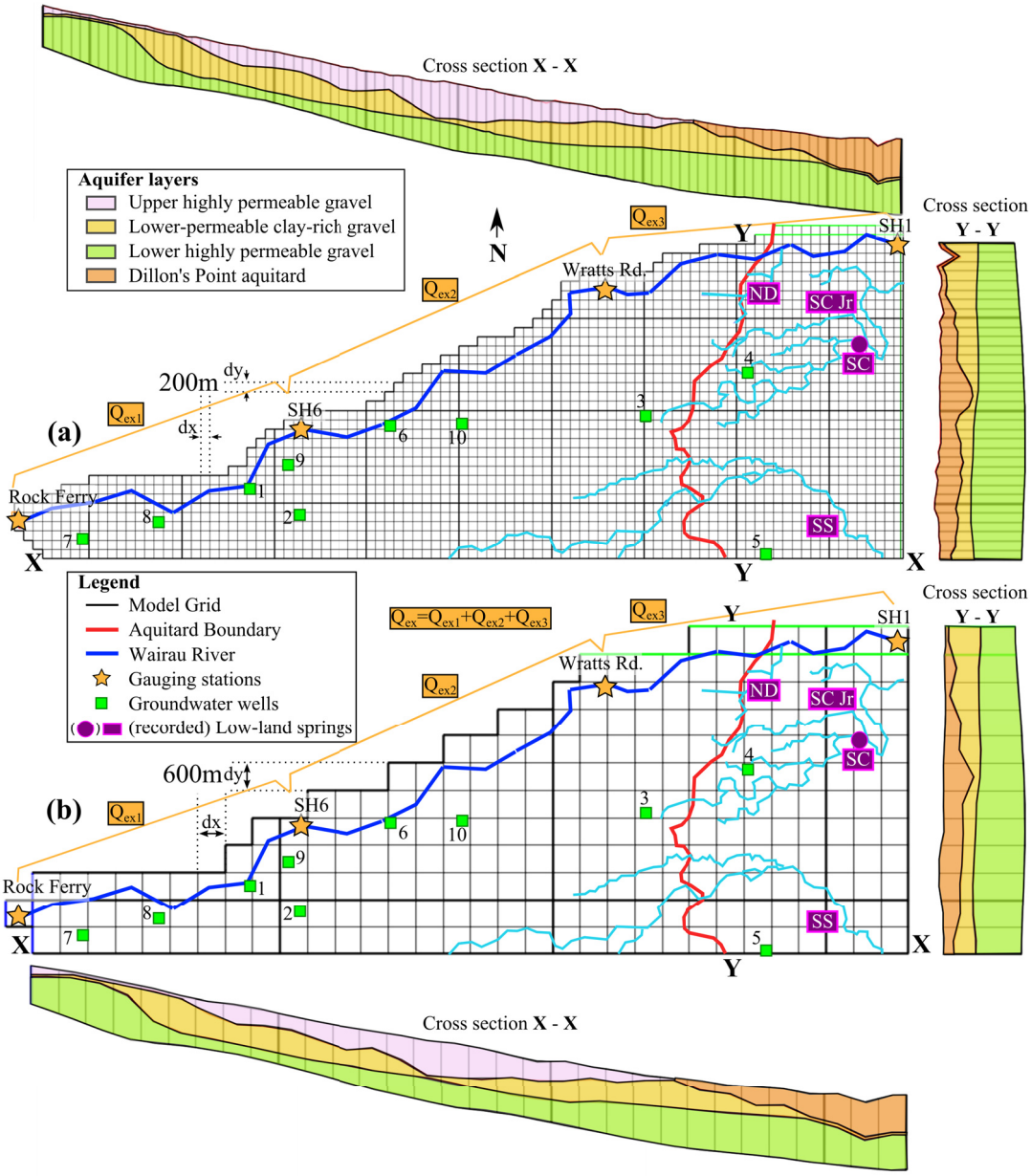

Figure 1: (a) Complex CM and (b) simple $SM_{1,sMm}$ MODFLOW models with the main geological features. Figure taken from Gosses and Wöhling (2019).

Table 2: Parameterization of the Wairau MODFLOW models (CM and $SM_{1,sMm}$). Table modified from Gosses and Wöhling (2019).

Parameter			**CM model**		**$SM_{1,sMm}$ model**	
Type [unit]	Property	Range	Type	Quantity	Type	Quantity
Hydraulic conductivity (K_H) [m/d]	Layer 1	$1 - 3,500$	Pilot Points	26	Zones	1
	Layer 2	$1 - 2,500$	Pilot Points	31	Zones	1
	Layer 3	$1 - 3,500$	Pilot Points	33	Zones	1
	Aquitard	$0.1 - 50$	Zones	1	Zones	1
Specific yield (S_y) [–]	Layer 1	$1 \times 10^{-3} - 0.35$	Pilot Points	26	Zones	1
	Layer 2	$1 \times 10^{-3} - 0.35$	Pilot Points	31	Zones	1
	Layer 3	$1 \times 10^{-3} - 0.35$	Pilot Points	33	Zones	1
	Aquitard	$1 \times 10^{-7} - 1 \times 10^{-3}$	Zones	1	Zones	1
Specific storage (S_S) [m^{-1}]	each	$1 \times 10^{-7} - 1 \times 10^{-3}$	Zones	4	Zones	4
Anisotropy ratio (f_a) [–]	each	$1 - 10$	Zones	4	Zones	4
River bed conductivity (K_R) [m/d]	–	$1 \times 10^{-3} - 0.2$	sections	12	sections	3
Drain bed conductivity (K_D) [m/d]	–	$1 \times 10^{-4} - 1,000$	sections	5	sections	5
Total				207		24

4.1.3 Surrogate 1: simplified MODFLOW model ($SM_{1,sMm}$)

The first of three different surrogate models belongs to the category of structural simplification methods and is a simplified version of the CM model also realized in MODFLOW, called $SM_{1,sMm}$. Most of the model specifications, boundaries and data are identical to the CM model, with the following differences:

- reduction of the number of active cells from 6,360 to 756 by universally coarsening of the horizontal grid from 200 $\times$ 200 m to 600 $\times$ 600 m,
- replacement of the spatially distributed K_H and S_y parameter fields by single parameter values per geological unit,
- and aggregation of the river bed sections, reducing the associated bed parameters from 12 to 3 and the total parameter number from 207 to 24.

The changes to the computational grid can be seen in Figure 1b, while Table 2 summarizes the changes in model parameters from the CM model to the $SM_{1,sMm}$ model.

4.1.4 Surrogate 2: linearized POD model ($SM_{2,POD}$)

Surrogate 2 is a POD model, and thus an application of the projection-based methods for surrogate modeling. It is based on a linearized version of the CM model I created in MATLAB© (Version R2015a) to allow application of the POD method named $SM_{2,POD}$. I linearized the CM model via

- using the river head values simulated by the stream-flow routing package (Niswonger et al., 2011) in the CM model along the Wairau River boundary to change the river boundary to a straight Cauchy boundary representation (identical to the river package of MODFLOW (Niswonger et al., 2011)),
- approximating the (originally unconfined) groundwater heads (for transmissivity, storage and boundary interaction) in the top-most model layer as the mean groundwater heads in each cell over the simulation period, and

- using the groundwater heads of $t-1$ for determining connected drain cells.

I conducted some preliminary studies to ensure that this linearized representation of the CM model reproduces simulated model outputs and uncertainty estimations with satisfactory accuracy.

I applied basic POD reduction to this linearized CM model. I created a snapshot set by running a full forward simulation and taking the simulated groundwater heads at every 50 days ($\mathbf{H}_s = [\mathbf{d}_{s1}; \mathbf{d}_{s51}; \mathbf{d}_{s101}; ...; \mathbf{d}_{s1221}; \mathbf{d}_{s1225}]$), and subsequently used singular value decomposition (SVD) to generate both singular values and left and right singular vectors. Following the methodologies common in literature, I chose a relative retained variance of 99.9 % to retain 7 left singular vectors for the subspace projection, and thus as the surrogate model dimension.

4.1.5 Surrogate 3: artificial neural networks ($\mathbf{SM_{3,ANN}}$)

Surrogate 3 used in this thesis are a suite of artificial neural networks (ANNs) designated $\mathrm{SM}_{3,\mathrm{ANN}}$. Artificial neural networks count as part of the data-driven methods, thus rounding out the three chosen surrogates composed of all three major surrogate methods introduced in section 2.1. For each data time series of interest, I created a single ANN model in the Neural Network Toolbox of MATLAB© (Version R2015a) following a basic configuration of a nonlinear autoregressive neural network with external input with parallel configuration (closed feedback loop). I optimized the number of nodes in the hidden layer, the amount of input delay and the amount of feedback delay for each data point beforehand. While time stepping was identical to the other models, I used river flow time series at Rock Ferry and total groundwater abstraction and recharge as forcings for the $\mathrm{SM}_{3,\mathrm{ANN}}$ models. I trained each of the 17 individual $\mathrm{SM}_{3,\mathrm{ANN}}$ models on the time series of its corresponding data type and location.

4.2 POD extension for explicit boundary treatment

POD is a popular surrogate modeling method in groundwater research that is not without problems, however. Thus, part of this thesis was to not only apply POD

as a surrogate model for the Wairau aquifer, but enhance the method in light of specific challenges of the test case. With a focus on surface water–groundwater interaction in the Wairau model, POD's erroneous handling of boundary conditions was chosen as a focus of my research in this regard. In this section, the extension for explicit boundary treatment to the proper orthogonal decomposition (POD) model reduction technique is presented. First, subsection 4.2.1 repeats the mathematical basis of (typical) groundwater models and the standard application of POD onto those models briefly. Then, the general form of the explicit boundary extension for POD (labeled eb-POD) is presented in subsection 4.2.2. I examine the impact of the three different types of boundary conditions on the application of eb-POD in subsection 4.2.3 and analyze the associated cost (in run time or accuracy) of eb-POD application in comparison to basic POD in subsection 4.2.4. Note that the methods presented in this section are reiterated in shortened form from Gosses et al. (2018).

4.2.1 Groundwater models and basic POD

POD works on the mathematical basis of physically based, spatially distributed groundwater models. Thus, I start by shortly summarizing the underlying principles. Most typical groundwater models like MODFLOW (Niswonger et al., 2011) are based on Darcy's law (Darcy, 1856) to solve a system of equations for groundwater heads at discretized time steps and locations. The equation system holds the hydrogeological conditions of the area of investigation in its definition of hydraulic conductivity, storage, sources/sinks and initial and boundary conditions. In its most basic form, a linear groundwater model with implicit time-stepping can be denoted in matrix form as:

$$\mathbf{A}\mathbf{d}^{t+1} = \mathbf{b}, \tag{1}$$

where the matrix $\mathbf{A}$ is a combination of stiffness matrix (containing the spatially distributed hydraulic conductivity) and mass matrix (containing the storage terms), $\mathbf{b}$ is a vector of boundary conditions and external forcings (such as groundwater recharge or pumping) and $\mathbf{d}$ is the vector of groundwater heads at

each discretized model location, or model cell. Note that this vector $\mathbf{d}$ is usually referred to as $\mathbf{h}$ in literature, but is renamed in this thesis to avoid overlap. The size of this equation system ($\mathbf{A}\epsilon\mathbb{R}^{n_m \times n_m}$ and $\mathbf{b}, \mathbf{d}\epsilon\mathbb{R}^{n_m \times 1}$) is directly dependent on on the number of model cells, n_m. It has to be solved for each time step to generate $\mathbf{d}$, which necessitates iterative manipulation of $\mathbf{A}$. This can become time-consuming for large models ($n_m \gg 1$), which is the foundation for the demand of model simplification.

POD obtains this reduction in size (and thus time) by projection of the equation system of Eq. (1) onto a suitable subspace of much smaller dimension. To generate a matrix capable of such subspace projection, the inherent variability of the complex model is captured in a matrix $\mathbf{H}_s$ of snapshots:

$$\mathbf{H}_s = [\mathbf{d}_{s1}, \mathbf{d}_{s2}, ..., \mathbf{d}_{sn_s}], \tag{2}$$

where each of the n_s vectors of $\mathbf{d}_s$ is one complex model solution for a certain time and forcing (i.e. a snapshot). This matrix $\mathbf{H}_s$ can be subjected to singular value decomposition, $\mathbf{H}_s = \mathbf{USV}^T$, where $\mathbf{U}\epsilon\mathbb{R}^{n_m \times n_m}$ is the matrix of left singular vectors, $\mathbf{S}\epsilon\mathbb{R}^{n_m \times n_s}$ is the diagonal matrix of singular values and $\mathbf{V}\epsilon\mathbb{R}^{n_s \times n_s}$ is the matrix of right singular vectors. Due to the nature of singular vectors (e.g. Strang, 2016), $\mathbf{U}$ holds the variability of $\mathbf{H}_s$ in descending order of importance as defined by the singular values in $\mathbf{S}$. This allows the definition of a projection matrix $\mathbf{P}\epsilon\mathbb{R}^{n_m \times n_r}$ as:

$$\mathbf{P} = [\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_{n_r}], \tag{3}$$

holding the n_r first columns $\mathbf{u}$ of $\mathbf{U}$. The cut-off n_r can be determined from the singular values in $\mathbf{S}$ by choosing a specific relative retained variance (RV %) of 99.99 %, for example). The projection matrix $\mathbf{P}$ defined by Eq. (3) is used to project the groundwater model as given in Eq. (1) onto a subspace via the Galerkin (1968) method:

$$\mathbf{P}^T\mathbf{APr}^{t+1} = \mathbf{P}^T\mathbf{b}. \tag{4}$$

As $n_r \ll n_m$, Eq. (4) can be solved for $\mathbf{r}$ much faster than Eq. (1) for $\mathbf{d}$. The groundwater heads in the original model discretization $\mathbf{h}$ are easily approximated from $\mathbf{r}$ via $\mathbf{d} \approxeq \mathbf{Pr}$.

4.2.2 Theory of explicit treatment of boundary conditions in POD

The general groundwater model as defined by Eq. (1) explicitly contains the boundary conditions of the model, enforcing the correct solution. In contrast, the simplified POD model of Eq. (4) only includes the boundary conditions implicitly inside its projection. This leads to errors of boundary groundwater heads and fluxes in the surrogate model. Here, I demonstrate a method of explicit boundary treatment in POD (eb-POD), based on research by Gunzburger et al. (2007), for all three types of boundary conditions in groundwater modeling: Dirichlet boundaries, Neumann boundaries and Cauchy boundaries.

First, I define the independent boundary conditions (bc) that I want to incorporate in the explicit treatment. Here, independent means a single model cell $bc \epsilon n_m$ for each forcing, even if that forcing affects multiple model cells (a constant groundwater head boundary along a certain model edge, for example). All bc rows of $\mathbf{P}$ are collected into a matrix $\mathbf{N} = \mathbf{P}(bc, \cdot)$, which is subsequently decomposed by QR-factorization into $\mathbf{Q}, \mathbf{R} = qr(\mathbf{N}^T)$ with $\mathbf{Q} \epsilon \mathbb{R}^{n_r \times n_r}$. If the number of independent boundary conditions, n_{bc}, is smaller than the size of the simplified model, n_r, the $n_{bc} + 1 : n_r$ columns of $\mathbf{Q}$ retain information of $\mathbf{P}$ independent of the boundaries bc. These $n_{bc} + 1 : n_r$ columns can now be used to compute:

$$\mathbf{\Psi} = \mathbf{PQ}(\cdot, n_{bc} + 1 : n_r), \tag{5}$$

where $\mathbf{\Psi} \epsilon \mathbb{R}^{n_{r2} \times n_{r2}}$ is a new projection matrix cleaned of the boundary information of bc and $n_{r2} = n_r - n_{bc}$. Combining Eq. (5) and Eq. (4) leads to:

$$\mathbf{\Psi}^T \mathbf{A} \mathbf{P} \mathbf{r}^{t+1} = \mathbf{\Psi}^T \mathbf{b}. \tag{6}$$

Eq. (6) is underdetermined as $\mathbf{\Psi}$ is smaller than $\mathbf{P}$ and $\mathbf{r}$ by n_{bc} and independent of the boundary conditions bc. Therefore, I add the rows containing

the explicit boundary conditions bc of the original model Eq. (1) in its simplified form, $\mathbf{A}(bc,\cdot)\mathbf{P}\mathbf{r}^{t+1} = \mathbf{b}(bc)$, to the new simplified model as defined by Eq. (6) to compute the simplified POD model with explicit treatment of the bc boundary conditions as:

$$\begin{bmatrix} \mathbf{\Psi}^T\mathbf{A}\mathbf{P} \\ \mathbf{A}(bc,\cdot)\mathbf{P} \end{bmatrix} \mathbf{r}^{t+1} = \begin{bmatrix} \mathbf{\Psi}^T\mathbf{b} \\ \mathbf{b}(bc) \end{bmatrix}. \tag{7}$$

From this simplified eb-POD model, I again approximate the groundwater heads via $\mathbf{d} \approxeq \mathbf{P}\mathbf{r}$.

4.2.3 Different boundary conditions in eb-POD

After summarizing the basic POD method and presenting the theory of the new eb-POD method, I now examine its impact dependent on type of boundary condition. Usually, three different types of boundary conditions are defined for groundwater modeling (e.g. Niswonger et al., 2011): Dirichlet boundaries (specified groundwater heads), Neumann boundaries (specified fluxes) and Cauchy boundaries (specified leakages). While the simplified eb-POD model as defined by Eq. (7) can be applied to all three boundary conditions, the effects or application differ slightly.

The system matrix $\mathbf{A}$ is very specific for Dirichlet boundaries (dc): $\mathbf{A}(dc,dc) = 1$, while all off-diagonal elements are 0. Thus, the dc rows in Eq. (7) can be slightly simplified, leading to an extended eb-POD equation of:

$$\begin{bmatrix} \mathbf{\Psi}^T\mathbf{A}\mathbf{P} \\ \mathbf{A}(bc,\cdot)\mathbf{P} \\ \mathbf{P}(dc,\cdot) \end{bmatrix} \mathbf{r}^{t+1} = \begin{bmatrix} \mathbf{\Psi}^T\mathbf{b} \\ \mathbf{b}(bc) \\ \mathbf{b}(dc) \end{bmatrix}, \tag{8}$$

with the unchanged approximation of $\mathbf{d} \approxeq \mathbf{P}\mathbf{r}$.

Neumann boundaries are simply treated explicitly via eb-POD as detailed in Eq. (7) (and thus, accordingly, via the second column of Eq. (8)).

Cauchy boundaries differ from Dirichlet and Neumann boundaries by being groundwater-head-dependent. While all boundary information for the other

types is contained in the vector **b**, Cauchy boundaries can be separated into a groundwater-head-independent part (added to the vector **b**, similar to Dirichlet and Neumann boundaries) and a groundwater-head-dependent part (added to the diagonal of **A** of the corresponding cell). This groundwater-head-dependent part is computed from the approximated groundwater heads in eb-POD ($\mathbf{Ad} \approxeq \mathbf{APr}$) and thus not error-free in the simplified model. The groundwater-head-independent part of the Cauchy boundary is treated explicitly in eb-POD (and thus error-free), similar to a Neumann boundary, though. This results in a reduction of the approximation error of the boundary flux in eb-POD compared to standard POD, but not an elimination of said error, as is shown in section 5.1 of this thesis.

4.2.4 Cost of eb-POD compared to basic POD

Above, I presented a method for explicit treatment of boundary conditions in POD, which enhances basic POD application in groundwater context by eliminating (for Dirichlet and Neumann boundaries) or reducing (for Cauchy boundaries) errors at boundary representation in the POD surrogate. In this subsection I demonstrate that this advantage of eb-POD comes at a cost, compared to basic POD, in either overall groundwater head field accuracy or surrogate model run time.

The exact implementation of the eb-POD method as presented above is one of two possible ones. In the shown case, the overall size of the simplified system does not change compared to the basic POD case ($n_r = const.$), but the size of the actual projection matrix is decreased ($n_{r2} = n_r - n_{bc}$). This means that the overall accuracy in estimating the groundwater head field of the eb-POD surrogate model is diminished, compared to a basic POD surrogate of the same size, at the cost of the explicit boundary treatment.

The other possible implementation would be to keep the size of the projection matrix constant ($n_{r2} = n_r$) by creating the new eb-POD projection matrix $\boldsymbol{\Psi}$ from a bigger original projection matrix $\mathbf{P} \epsilon \mathbb{R}^{n_m \times n_{rb}}$ with $n_{rb} = n_r + n_{bc}$. This ensures a preservation of the overall accuracy of the surrogate model, but increases the size

of the eb-POD surrogate compared to the basic POD surrogate ($n_{rb} > n_r$). This size increase is equivalent to an increase in run time of the surrogate model.

The explicit treatment of boundary conditions with eb-POD is thus always accompanied by an additional cost in comparison to the basic POD method: either decrease of overall groundwater head accuracy or increase of model run time. This is further evaluated in section 5.1.

4.3 Model simplification error analysis – theory

After introduction of the research area, chosen benchmark and surrogate models and the enhancement of boundary representation for POD, I next introduce the concept of model simplification error. I previously defined the simplification error as the discrepancy between complex and surrogate model outputs, and give a more mathematical definition of the simplification error in this section. The applied methodology to estimate the simplification error and its contributions via scatter plots using a paired model approach follows in section 4.4. In this theory section, I start out by defining a linear model and its solution space and null-space mathematically in subsection 4.3.1. Then, I characterize the surrogate model and its calibration in subsection 4.3.2. Subsection 4.3.3 holds the mathematical relation of complex model and surrogate model parameters for the analysis of parameter surrogacy. Finally, I derive the theoretical definition of the simplification error itself in subsection 4.3.4. The concepts presented here are based on the assumption of model linearization. While this should not prevent its use for non-linear groundwater models, it can affect the results (and is pointed out in the discussion, where necessary). Note that this section is an abbreviated repetition of the work published by Gosses and Wöhling (2019), which itself is based on the work by Doherty and Christensen (2011) and Watson et al. (2013).

4.3.1 A linear model, solution space and null-space

As a basis for the mathematical definition of the surrogate model, parameter simplification and simplification error, I start with some necessary definitions. Notation-wise, the model presented in this subsection represents the complex

benchmark model (but the presented concepts are of general nature). A linear model can be described as

$$\mathbf{h} = \mathbf{Z}\mathbf{k}+\boldsymbol{\varepsilon}, \tag{9}$$

where $\mathbf{h}$ denotes the calibration data set, $\mathbf{Z}$ is a matrix representing the model's action under calibration, the vector $\mathbf{k}$ holds the model parameters and $\boldsymbol{\varepsilon}$ is a vector of (random) measurement noise. The null-space of $\mathbf{Z}$ is defined as a non-zero solution for $\mathbf{k}_n$ that results in $\mathbf{0} = \mathbf{Z}\mathbf{k}_n$. Combining this null-space definition with Eq. (9) leads to

$$\mathbf{h} = \mathbf{Z}(\underline{\mathbf{k}} + \mathbf{k}_n) + \boldsymbol{\varepsilon}, \tag{10}$$

where $\underline{\mathbf{k}}$ is a vector of calibrated parameter values. Parameter non-uniqueness is present where multiple non-zero vectors $\mathbf{k}_n$ exist. Eq. (10) shows the contributions of solution space ($\underline{\mathbf{k}}$, which can be inferred through calibration) and null-space ($\mathbf{k}_n$, which can not be inferred) to the model parameters.

4.3.2 Surrogate model: definition and calibration

After defining the complex model above, I denote the mathematical representation of surrogate models in this subsection. In the context of the linear paired model analysis, model simplification is defined as omission of complex model parameters. Conceptually, this can also include simplification and/or omission of processes. One can partition $\mathbf{k} = \begin{bmatrix} \mathbf{k}_s \\ \mathbf{k}_o \end{bmatrix}$ into surrogate model parameters $\mathbf{k}_s$ and omitted parameters $\mathbf{k}_o$. Combining this with Eq. (9) leads to

$$\mathbf{h} = \mathbf{Z}_s\mathbf{k}_s + \mathbf{Z}_o\mathbf{k}_o + \boldsymbol{\varepsilon}, \tag{11}$$

where $\mathbf{Z}_s$ and $\mathbf{Z}_o$ are segments of the system matrix $\mathbf{Z}$ related to $\mathbf{k}_s$ and $\mathbf{k}_o$. Let us undertake singular value decomposition (SVD) of the surrogate model.

The SVD of $\mathbf{Z}_s$ is defined as $\mathbf{Z}_s = \mathbf{USV}^T$, where $\mathbf{S}$ holds the singular values of $\mathbf{Z}_s$ and $\mathbf{U}$ and $\mathbf{V}$ the left- and right-singular vectors. Under the assumption that the surrogate model has a null-space as well, $\mathbf{V}$, $\mathbf{S}$ and $\mathbf{U}$ can be partitioned into their solution space ($\mathbf{V}_1$, $\mathbf{S}_1$ and $\mathbf{U}_1$) and null-space ($\mathbf{V}_2$, $\mathbf{S}_2$ and $\mathbf{U}_2$) counterparts. Now, the calibrated surrogate model parameter set is written as

$$\underline{\mathbf{k}}_s = \mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^T\mathbf{h} - \mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^T\boldsymbol{\varepsilon}. \tag{12}$$

4.3.3 Parameter simplification – relationship between complex model and surrogate model parameters

The definitions of complex model and surrogate model in the previous subsections can be used to analyze the relationship between their respective parameters. Where the parameters in the surrogate model differ from the complex model parameters, process detail gets lost. This can result in parameter surrogacy, meaning that surrogate model parameters are made up of different complex model parameters. This leads to a disconnection between real-world processes and parameter representation for surrogate model parameters that are supposed to describe such processes. In such cases, surrogate model parameters such as hydraulic conductivities can no longer be informed by geological knowledge (by subjecting them to geology-derived limits, for example), but must instead be treated as "effective" parameters, devoid of real-world process representation. Below, this concept of parameter surrogacy is demonstrated mathematically.

The complex model parameters can be written as:

$$\underline{\mathbf{k}} = \begin{bmatrix} \underline{\mathbf{k}}_s \\ \mathbf{k}_o \end{bmatrix} = \begin{bmatrix} \mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^T \\ \mathbf{0} \end{bmatrix} \mathbf{h} - \begin{bmatrix} \mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^T \\ \mathbf{0} \end{bmatrix} \boldsymbol{\varepsilon}, \tag{13}$$

based on Eq. (12). I combine Eq. (13) with Eq. (11) to arrive at:

$$\underline{\mathbf{k}} = \begin{bmatrix} \underline{\mathbf{k}}_s \\ \mathbf{k}_o \end{bmatrix} = \begin{bmatrix} \mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^T \\ \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{Z}_s & \mathbf{Z}_o \end{bmatrix} \begin{bmatrix} \mathbf{k}_s \\ \mathbf{k}_o \end{bmatrix} - \begin{bmatrix} \mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^T \\ \mathbf{0} \end{bmatrix} \boldsymbol{\varepsilon}. \tag{14}$$

Ignoring measurement noise for ease of understanding, Eq. (14) can be reorganized to solve for the calibrated surrogate model parameter set $\underline{\mathbf{k}}_s$ and combined with the SVD of $\mathbf{Z}_s = \mathbf{USV}^T$ to gain:

$$\underline{\mathbf{k}}_s = \mathbf{V}_1\mathbf{V}_1^T\mathbf{k}_s + \mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^T\mathbf{Z}_o\mathbf{k}_o. \tag{15}$$

Eq. (15) shows that the calibrated surrogate parameter set is made up of two parts. The first, $\mathbf{V}_1\mathbf{V}_1^T\mathbf{k}_s$, represents the projection of the surrogate model parameter set onto its solution space. The second part contains $\mathbf{k}_o$, the omitted complex parameter set. This represents the parameter surrogacy, where the surrogate model parameter set is partially representing omitted complex parameter detail.

This process can be further analyzed by representation of the surrogate model parameters trough their composition in respect to complex model parameters. Fitting the surrogate model parameter set, $\mathbf{k}_s$, to output of the complex model, $\mathbf{o}$, can be written as:

$$\mathbf{k}_s = (\mathbf{Z}_s^T\mathbf{Z}_s)^{-1}\mathbf{Z}_s^T\mathbf{o}. \tag{16}$$

Adapt Eq. (9) by substituting $\mathbf{h}$ by $\mathbf{o}$ and combine it with Eq. (16), while ignoring measurement noise to get:

$$\mathbf{k}_\mathrm{s} = (\mathbf{Z}_\mathrm{s}^T\mathbf{Z}_\mathrm{s})^{-1}\mathbf{Z}_\mathrm{s}^T\mathbf{Z}\mathbf{k} = \mathbf{L}\mathbf{k}. \tag{17}$$

Eq. (17) relates the surrogate model parameter set, $\mathbf{k}_s$, to the complex model parameter set, $\mathbf{k}$, via the matrix $\mathbf{L}$. This relationship is used in this thesis to

compute the composition of surrogate model parameters in regard of complex model parameters to reveal the existence of parameter surrogacy.

4.3.4 Simplification error of surrogate model predictions

While the subsection above compared complex and surrogate models in regard to their parameters, the main focus of this section is the definition of the model simplification error. As stated earlier, I define the model simplification error as the discrepancy between complex model and surrogate model output due to model simplification. Generally, a single model output of interest, for which the model simplification error is estimated, can be defined as a model prediction. Mathematically, a model prediction is defined as

$$s = \mathbf{y}^T \mathbf{k}, \tag{18}$$

where scalar s is the prediction and the vector $\mathbf{y}$ holds the sensitivities of the prediction to the model parameters $\mathbf{k}$. $\mathbf{k}$ can be separated into surrogate model parameters and omitted parameters, leading to $s = \mathbf{y}_s^T \mathbf{k}_s + \mathbf{y}_o^T \mathbf{k}_o$. A prediction made by the calibrated surrogate model would be given as $\underline{s} = \mathbf{y}_s^T \underline{\mathbf{k}}_s$, leading to a definition of the error between surrogate and complex model prediction as

$$\underline{s} - s = \mathbf{y}_s^T \underline{\mathbf{k}}_s - \mathbf{y}_s^T \mathbf{k}_s - \mathbf{y}_o^T \mathbf{k}_o. \tag{19}$$

Combining Eq. (19) with Eq. (12) (ignoring measurement error for clarity) leads to $\underline{s} - s = \mathbf{y}_s^T \mathbf{V}_1 \mathbf{S}_1^{-1} \mathbf{U}_1^T \mathbf{h} - \mathbf{y}_s^T \mathbf{k}_s - \mathbf{y}_o^T \mathbf{k}_o$. Combining this with Eq. (11) and further simplification and rearranging (see Doherty and Christensen (2011) for details) leads to the final definition of the model simplification error as

$$s = \underbrace{\underline{s}}_{\text{term 0}} + \underbrace{\mathbf{y}_s^T \mathbf{V}_2 \mathbf{V}_2^T \mathbf{k}_s}_{\text{term 1}} + \underbrace{\mathbf{y}_s^T \mathbf{V}_1 \mathbf{S}_1^{-1} \mathbf{U}_1^T \boldsymbol{\varepsilon}}_{\text{term 2}} - \underbrace{\mathbf{y}_s^T \mathbf{V}_1 \mathbf{S}_1^{-1} \mathbf{U}_1^T \mathbf{Z}_o \mathbf{k}_o}_{\text{term 3}} + \underbrace{\mathbf{y}_o^T \mathbf{k}_o}_{\text{term 4}}. \tag{20}$$

Eq. (20) states the complex model prediction s as a combination of the surrogate model prediction $\underline{s}$ (term 0) and four different contributions (term 1–4) to the simplification error of this prediction. I now present the applied methodology of paired model analysis to estimate this simplification error and its contributions.

4.4 Model simplification error analysis – scatter plot analysis

After presenting the theoretical background of complex model, surrogate model and the simplification error in section 4.3, this section holds the methodology for the estimation of the simplification error via paired model analysis generating scatter plots of complex model–surrogate model predictions. First, I present the actual process of creating the scatter plots, along with introducing the necessary symbols and a schematic example, in subsection 4.4.1. As the resulting scatter plots hold a variety of information, subsections 4.4.2 and 4.4.3 deal with the general features of the plot and the contributions of the different error terms to the overall simplification error (see above), respectively. In subsection 4.4.4, I present different prediction pairs which can be added to the scatter plots for further information gain. Finally, I summarize the main insights about simplification error gained from the scatter plot analysis in subsection 4.4.5.

4.4.1 Methodology

The model simplification error and its contributions presented in Eq. (20) are best analyzed via scatter plots of paired model predictions. In the following, the methodology to create those scatter plots is detailed in 11 steps, taken from Gosses and Wöhling (2019) with slight alterations. The analysis of the plots themselves is presented in the subsequent subsections.

1. Use all available system knowledge to create a complex model $\mathbf{Z}$ which incorporates any processes necessary. Its parameter set is $\mathbf{k}$.

2. Run the model for n different model parameter realizations $\mathbf{k}_i$ (for $i = 1, \ldots, n$) and record the accompanying model output $\mathbf{o}_i$. As these realiza-

tions $\mathbf{k}_i$ represent the unknown variability of the different parameters, they should be random, with the statistics (such as type of distribution, mean and standard deviation etc.) informed by the data or expert knowledge about the system.

3. Build a surrogate model $\mathbf{Z}_s$ that allows for stochastic calibration (i.e. runs fast).

4. Corrupt the complex model outputs with applicable measurement errors, $\mathbf{o}_i + \boldsymbol{\varepsilon}_i$, and calibrate the surrogate model against each of those corrupted outputs. This results in a calibrated surrogate model realization for each random complex model realization.

5. Compute complex model predictions $\mathbf{s}_i^j$ and calibrated surrogate model predictions $\underline{s}_i^j$ for each realization and all $j = 1, \ldots, m$ predictions.

6. Plot the surrogate model predictions $\underline{s}_i^j$ versus complex model predictions s_i^j scatter points for each prediction j and fit a linear regression line $s = a\underline{s} + b$ to the $[s_i; \underline{s}_i]$ pairs. Add 95 % confidence intervals (which are $\pm 2\sigma$ assuming normal distribution) to the plots. The confidence intervals are used to discern the contributions to predictive uncertainty by the four terms of Eq. (20). This is shown in Figure 2a and is explained subsequently.

7. Calibrate the complex model $\mathbf{Z}$ against the real-world measurement data set $\mathbf{h}$, resulting in a complex model output $\mathbf{o}(\mathbf{h})$ and calibrated complex model predictions $s_{\mathbf{h}}$.

8. Use the calibrated complex model output $\mathbf{o}(\mathbf{h})$ as calibration targets for the surrogate model $\mathbf{Z}_s$, resulting in corresponding surrogate model predictions $\underline{s}_{\mathbf{o}(\mathbf{h})}$. The $[\underline{s}_{\mathbf{o}(\mathbf{h})}; s_{\mathbf{h}}]$ pairs of (complex model-) calibrated surrogate model predictions and calibrated complex model predictions are the basis for the simplification error analysis and bias correction which is explained subsequently.

9. The predictions $\underline{s}_{\mathbf{o(h)}}$ of the surrogate model (calibrated against the output $\mathbf{o}(\mathbf{h})$ of the real-world-calibrated complex model) are projected via the regression line from step 6 into corresponding surrogate model predictions $\hat{\underline{s}}_{\mathbf{o(h)}}$ (corrected for bias by model simplification). This is shown in Figure 2b.

10. Calibrate the surrogate model $\mathbf{Z}_s$ against the real-world measurement data set $\mathbf{h}$ and compute the corresponding model predictions $\underline{s}_{\mathbf{h}}$.

11. If available, add the data pairs of $[\underline{s}^j_{\mathbf{h}}; h^j]$ to the plot, where h is real-world data corresponding to the prediction. $\boldsymbol{\varepsilon}_{CM} = s^j_{\mathbf{h}} - h^j$ is the discrepancy of the corresponding j-th complex model prediction and reality, as shown schematically in Figure 2c.

4.4.2 General features of the scatter plots

As the scatter plots created via the methodology above are the main tool for the analysis of the simplification error as defined in Eq. (20) and hold a multitude of information, I provide some context on reading them, starting with some general features.

Assuming that the surrogate model is a perfect representation of the complex model ($\mathbf{Z}_s = \mathbf{Z}$ and $\mathbf{k}_s = \mathbf{k}$), Eq. (20) would simplify to $s = \underline{s}$. This would result in a scatter plot of points all one a line of slope 1 with intercept 0, meaning the different contributions to the model simplification error manifest as deviations of the scatter points from this line. Figure 2a shows an exemplary scatter plot with regression line and 95 % uncertainty interval (steps 1–6). The regression line is defined by slope a and intercept b. The intercept b is the (mean) bias of the surrogate model prediction compared to the complex model prediction. A slope of $a < 1$ indicates more horizontal than vertical scatter. As this means that calibrated surrogate model predictions are more variable than random complex model predictions, it suggests that random (uncalibrated) complex model predictions actually have a smaller propensity for error than calibrated surrogate model

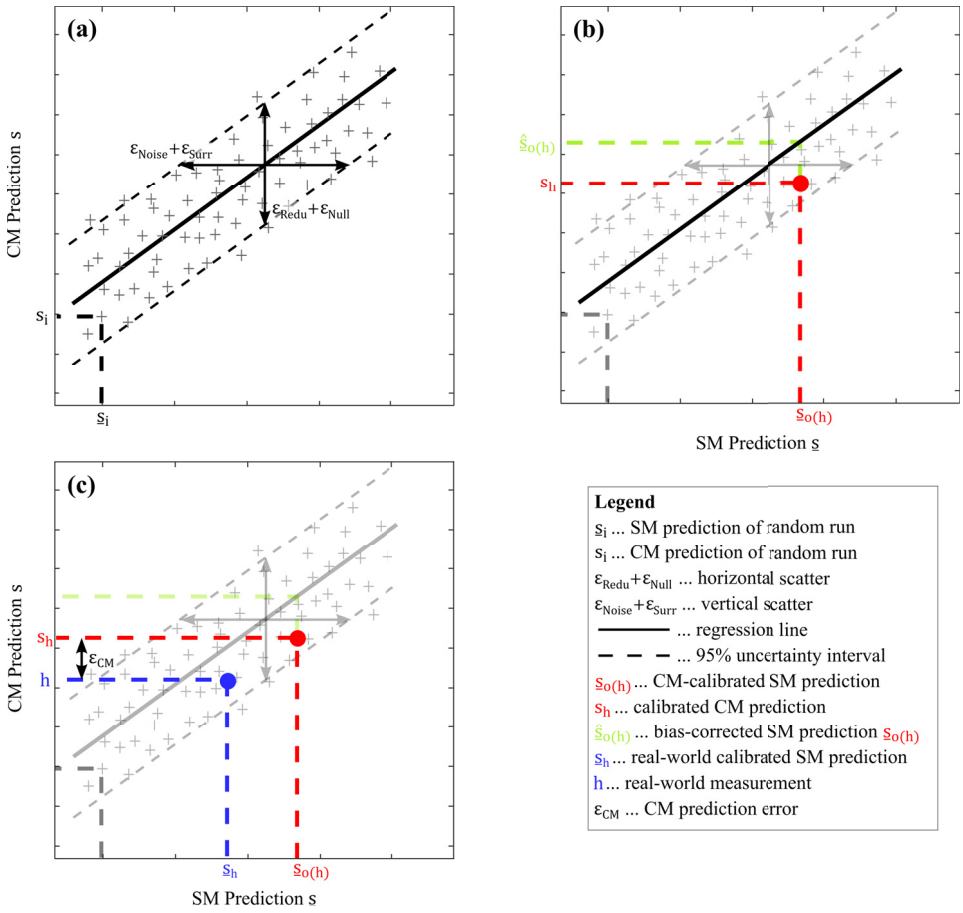

Figure 2: Schematic of a paired model analysis scatter plot. (a) shows the scatter points, the regression line, its uncertainty intervals and the horizontal and vertical scatter contributions. (b) adds the bias correction of the (CM calibrated) SM prediction. (c) shows the CM model prediction error. Figure modified from Gosses and Wöhling (2019).

predictions. Analogously, a slope of $a > 1$ shows more variability in random complex model predictions than in surrogate model predictions. This identifies an over-fitting in the calibration or a missing process detail in the surrogate model obstructing it to reproduce the full range of random complex model predictions.

4.4.3 Contributions of error terms

Besides the information of slope and intercept presented above, the scatter plots created by the above methodology (as shown in Figure 2a) also allow distinguishing the model simplification error and its four error terms (1–4) defined in the theoretical characterization of the model simplification error in Eq. (20). As evident from Eq. (20), an omission of all four error terms would result in $s = \underline{s}$, and thus a perfect situation of all points on a line of slope 1 with intercept 0. Thus, the scatter holds information about the four different error terms:

Term 1 of Eq. (20), $\mathbf{y}_s^T \mathbf{V}_2 \mathbf{V}_2^T \mathbf{k}_s$, represents the null-space component of the simplification error and is thus labeled $\boldsymbol{\varepsilon}_{Null}$. As the null-space (of the surrogate model) is, by definition, insensitive to the surrogate model, it does not affect $\underline{s}$ and therefore results in vertical scatter in the plot. Term 2 of Eq. (20), $\mathbf{y}_s^T \mathbf{V}_1 \mathbf{S}_1^{-1} \mathbf{U}_1^T \boldsymbol{\varepsilon}$, signifies the effect of measurement noise. As measurement noise is added onto the complex model outputs before calibration of the surrogate model in step 4 of the above methodology, it only influences $\underline{s}$. Therefore, it adds horizontal scatter labeled as $\boldsymbol{\varepsilon}_{Noise}$. So does the effect of term 3 of Eq. (20), $\mathbf{y}_s^T \mathbf{V}_1 \mathbf{S}_1^{-1} \mathbf{U}_1^T \mathbf{Z}_o \mathbf{k}_o$. This represents the effects of parameter surrogacy, labeled $\boldsymbol{\varepsilon}_{Surr}$, onto the surrogate model. This concept, which describes a contribution of complex model parameters and processes onto surrogate model parameters, is explained in more detail below. Term 4 of Eq. (20), $\mathbf{y}_o^T \mathbf{k}_o$, is the contribution of complex parameter omission onto the simplification error, labeled ε_{Redu}. Omitted parameters can only influence s (and not $\underline{s}$), adding vertical scatter.

4.4.4 Prediction pairs

Thus far, analysis of the scatter plots (in regard to surrogate model bias, the simplification error and its different contributions) was limited to the scatter points,

the regression line and its uncertainty intervals as presented in Figure 2a and Steps 1–6 of the methodology given in subsection 4.4.1. Steps 7–11 of the above methodology create several special prediction pairs (shown in Figure 2b&c) which allow further insight into the simplification error of the surrogate for this model prediction.

Figure 2b denotes the first two prediction pairs, both based on $\underline{s}_{\mathbf{o(h)}}$. $\underline{s}_{\mathbf{o(h)}}$ is the prediction of the surrogate model, calibrated against $\mathbf{o(h)}$, the output of the complex model calibrated against the real-world data set $\mathbf{h}$.

The $[\underline{s}_{\mathbf{o(h)}}; s_{\mathbf{h}}]$ pair shows the intercept between $\underline{s}_{\mathbf{o(h)}}$ and its counterpart $s_{\mathbf{h}}$, the prediction of the complex model calibrated against the real-world data set $\mathbf{h}$. The extent of vertical scatter around this $[\underline{s}_{\mathbf{o(h)}}; s_{\mathbf{h}}]$ intercept is a graphical representation of the simplification error. Thus, the $[\underline{s}_{\mathbf{o(h)}}; s_{\mathbf{h}}]$ pair should lie within the uncertainty bands most (i.e. 95 %) of the time. Therefore, the $[\underline{s}_{\mathbf{o(h)}}; s_{\mathbf{h}}]$ pair is a test of the applicability of the method for the specific prediction.

As noted above, the intercept b of the regression line gives the (mean) bias of the surrogate model predictions. Thus, we can use the regression line to correct $\underline{s}_{\mathbf{o(h)}}$ for this bias, resulting in $\hat{\underline{s}}_{\mathbf{o(h)}}$.

In some cases, predictions are analyzed that are reproductions of measurement data sets, like groundwater head data during a validation period, for example. In such cases where real-world data corresponding to the prediction, h^j, is available, another prediction pair can be added to the graph. Figure 2c shows the addition of the $[\underline{s}_{\mathbf{h}}; h]$ pair to the scatter plot. $\underline{s}_{\mathbf{h}}$ denotes the prediction of the surrogate model calibrated against the real-world data set $\mathbf{h}$, and the real-world "prediction", h, is added to the y-axis. A complex model perfectly reproducing reality would result in $h = s_{\mathbf{h}}$ (ignoring measurement error in h). Where this is not the case, we can draw $\boldsymbol{\varepsilon}_{CM} = s_{\mathbf{h}} - h$, depicting an estimate of the prediction-specific error of the calibrated model against reality.

4.4.5 Summary

In summary, the following findings can be gained from the paired complex model surrogate model scatter plots:

- The intercept b of the regression line denotes the mean bias of the surrogate model prediction and can thus be used for bias-correction of $\underline{s}_{\mathbf{o(h)}}$ to $\hat{\underline{s}}_{\mathbf{o(h)}}$.
- The slope a of the regression line hints at potential over-fitting ($a > 1$) or reveals where the propensity for predictive error actually increased through surrogate model calibration ($a < 1$).
- Influence of measurement noise and parameter surrogacy (terms 2 and 3 of Eq. (20)) are given as horizontal scatter, while null-space contribution and structural simplification (terms 1 and 4 of Eq. (20)) create the vertical scatter. The vertical scatter denotes the simplification error.
- Where the $[\underline{s}_{\mathbf{o(h)}}; s_{\mathbf{h}}]$ pair lies outside the uncertainty bands, it reveals an underestimation of the simplification error for this prediction.
- Available real-world validation data corresponding to the prediction, h, allows the estimation of the prediction-specific error of the complex model, ε_{CM}.

4.5 Robust data worth analysis

In this section, I present the robust data worth analysis applied in this thesis. As illustrated in section 2.2, first-order second-moment uncertainty and data worth estimation methods are dependent on the current calibrated parameter set. Thus, their results estimates do not incorporate the non-uniqueness of calibrated parameter values and could therefore be highly erroneous. As presented in section 2.2.4, stochastic uncertainty and data worth estimation is highly computationally expensive, though. In this thesis I propose to instead combine first-order second-moment uncertainty and data worth estimation methods with a highly efficient method to generate calibrated model parameter sets to a more robust data worth

analysis method. First, I summarize the first-order second-moment (predictive) uncertainty estimation methods developed by Doherty (2016) (4.5.1). Then, I present how to apply these methods to compute the worth of three different categories of data (existing, "future" and "parametric") in regards to their effect on predictive postcalibration uncertainty (4.5.2). I apply these first-order second-moment uncertainty and data worth estimation techniques in a more robust data worth analysis to account for parametric uncertainty. For this, I present the methodology for efficient generation of calibrated parameter set via null-space perturbation, again taken from Doherty (2016), in abbreviated form next (4.5.3). Finally, I combine the first-order second-moment uncertainty and data worth estimation methods with the efficient parameter set generation into the robust data worth analysis applied in this thesis (4.5.4).

4.5.1 First-order second-moment uncertainty estimation

As stated above, I utilize the methodology and tools presented by Doherty (2016) to characterize uncertainty (and changes to it due to different types of data) of model predictions at a linearized model state (i.e. for one calibrated parameter set). Starting with the definition of a model prediction as given in Eq. (18), one can describe the uncertainty of the prediction by its variance σ_s^2 given as:

$$\sigma_s^2 = \mathbf{y}^T \mathbf{C}(\mathbf{k}) \mathbf{y}, \tag{21}$$

where $\mathbf{C}(\mathbf{k})$ is the covariance matrix of the parameters. Note that this is an estimation of the precalibration uncertainty, as it is not subjected to data. To compute the postcalibration uncertainty from this term, I start by combining Eq. (18) (the definition of a model prediction) and Eq. (9) (the definition of the linear model) to get:

$$\begin{bmatrix} s \\ \mathbf{h} \end{bmatrix} = \begin{bmatrix} \mathbf{y}^T & 0 \\ \mathbf{Z} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{k} \\ \boldsymbol{\varepsilon} \end{bmatrix}, \tag{22}$$

where $\mathbf{I}$ is the identity matrix. The covariance of Eq. (22) is thus defined as:

$$\mathbf{C}\left(\begin{bmatrix} s \\ \mathbf{h} \end{bmatrix}\right) = \begin{bmatrix} \mathbf{y}^T\mathbf{C}(\mathbf{k})\mathbf{y} & \mathbf{y}^T\mathbf{C}(\mathbf{k})\mathbf{Z}^T \\ \mathbf{Z}\mathbf{C}(\mathbf{k})\mathbf{y} & \mathbf{Z}\mathbf{C}(\mathbf{k})\mathbf{Z}^T + \mathbf{C}(\boldsymbol{\varepsilon}) \end{bmatrix}. \tag{23}$$

For an exemplary vector $\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}$ (analogous to Eq. (22)), the covariance matrix is $\mathbf{C}(\mathbf{x}) = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix}$ (analogous to Eq. (23)). From this, the conditional variance is defined as $\mathbf{C}_{1|2} = \mathbf{C}_{11} - \mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{21}$. Now, this definition of the conditional variance is used analogous for Eq. (23), which gives the computation of $C_{s|\mathbf{h}} \equiv \sigma^2_{s|\mathbf{h}}$, the variance of the prediction s conditional of the calibration data $\mathbf{h}$, as:

$$\sigma^2_{s|\mathbf{h}} = \mathbf{y}^T\mathbf{C}(\mathbf{k})\mathbf{y} - \mathbf{y}^T\mathbf{C}(\mathbf{k})\mathbf{Z}^T[\mathbf{Z}\mathbf{C}(\mathbf{k})\mathbf{Z}^T + \mathbf{C}(\boldsymbol{\varepsilon})]^{-1}\mathbf{Z}\mathbf{C}(\mathbf{k})\mathbf{y}. \tag{24}$$

Note how this definition of the predictive postcalibration uncertainty is computed as a reduction from the precalibration uncertainty ($\mathbf{y}^T\mathbf{C}(\mathbf{k})\mathbf{y}$, as defined in Eq. (21)) through the calibration with the data $\mathbf{h}$.

Computing $\sigma^2_{s|\mathbf{h}}$ not only necessitates $\mathbf{Z}$ (the Jacobian matrix of the calibrated model, which can easily be calculated) and $\mathbf{y}$ (the column of $\mathbf{Z}$ pertaining to the prediction s of interest), but also the covariance matrices of the parameters ($\mathbf{C}(\mathbf{k})$) and observations ($\mathbf{C}(\boldsymbol{\varepsilon})$), respectively. Where spatially distributed, correlated parameters (like pilot points of hydraulic conductivity) are used, PEST (Doherty, 2016) provides tools for the computation of the accompanying parts of the $\mathbf{C}(\mathbf{k})$ matrix. For non-correlated parameters, I fill the diagonal of $\mathbf{C}(\mathbf{k})$ with estimations of the variance of the respective parameter. Observations are usually thought to be uncorrelated, so $\mathbf{C}(\boldsymbol{\varepsilon})$ is defined as a diagonal matrix proportional to the estimated measurement errors of the observations. Note that the model can include a multitude of "observations" that do not have actual measurements associated to them (model predictions or similar) or are not used in the calibration process. These have diagonal entries of 0 in the $\mathbf{C}(\boldsymbol{\varepsilon})$ matrix. Throughout this thesis, predictive uncertainty refers to the predictive postcalibration uncertainty as defined by Eq. (24).

4.5.2 Worth of data

The effects of different data on the predictive uncertainty as defined by Eq. (24) can be classified into three different categories depending on the type of data:

1. existing data: increase of predictive uncertainty through omission of existing measurement data,

2. "future" data: decrease of predictive uncertainty through inclusion of "future" measurement data, where "future" means not part of the existing measurement data set, and can thus include future, non-existing measurements, and

3. "parametric" data: decrease of predictive uncertainty through (assumed) perfect knowledge of model parameters

The worth of all three data types in regard to changing predictive uncertainty can be estimated through the use of the methodology applied above.

Subtraction of existing data and the potential increase in predictive uncertainty is simply computed by setting the diagonal entry of $\mathbf{C}(\boldsymbol{\varepsilon})$ of the observation of choice to 0 and comparing the resulting $\sigma^2_{s|\mathbf{h}}$ to the base case. To estimate the uncertainty reduction through addition of "future" data, the chosen measurement types, locations and times are defined as measurements in the model software, are added to the observation data set and the Jacobian matrix is recalculated (to gain $\mathbf{Z}$ and $\mathbf{y}$ with entries for this "future" data). Then, the "future" data can be added to the uncertainty estimation process by assigning them non-zero diagonal entries in the $\mathbf{C}(\boldsymbol{\varepsilon})$ matrix (again in proportionality to their estimated measurement error). A (potential) reduction in predictive uncertainty through the inclusion of this "future" data is then compared to the base case. It is important to emphasize that no actual measured data for such "future" data is necessary for these computations. Hence, the method can be used to preemptively estimate the worth of "future" data in reducing the postcalibration uncertainty of model predictions.

A similar approach is taken to estimate the worth of "parametric" data on predictive uncertainty. By setting the variance of the parameter in question to 0

in $\mathbf{C}(\mathbf{k})$, the model assumes perfect knowledge of this parameter. The resulting (potential) reduction in $\sigma^2_{s|\mathbf{h}}$ can again be compared to the base case as an estimate of the worth of this "parametric" data for postcalibration uncertainty of the chosen model predictions.

4.5.3 Generating calibrated parameter sets – null-space parameter perturbation

The data worth estimation methods presented above are linear in respect to the calibrated parameter set. Accounting for parameter non-uniqueness demands multiple parameter sets. I present an efficient generation of calibrated parameters sets using tools and methods provided by the PEST (Doherty, 2016) toolbox, where they are called null-space Monte-Carlo techniques.

After the original calibration exercise which leads to one (of potentially many) calibrated parameter set $\underline{\mathbf{k}}$, it consists of the following steps:

1. Creation of i random parameter sets $\mathbf{k}_i$ based on the prior parameter probability distribution as defined by $\mathbf{C}(\mathbf{k})$.

2. The original calibrated parameter set $\underline{\mathbf{k}}$ is subtracted from the random parameter sets $\mathbf{k}_i$, yielding i parameter difference sets $\mathbf{k}_i - \underline{\mathbf{k}}$.

3. Projection of the parameter difference sets $\mathbf{k}_i - \underline{\mathbf{k}}$ onto the null-space of $\underline{\mathbf{k}}$ to generate $(\mathbf{k}_i - \underline{\mathbf{k}})^{ns}$.

4. Addition of the null-space projected parameter difference sets $(\mathbf{k}_i - \underline{\mathbf{k}})^{ns}$ onto the original calibrated parameter set, $\underline{\mathbf{k}} + (\mathbf{k}_i - \underline{\mathbf{k}})^{ns}$.

For a linear model, this process would result in different parameter sets that all calibrate the model at the same precision as the original calibrated parameter set $\underline{\mathbf{k}}$, as they only differ from $\underline{\mathbf{k}}$ through components in its null-space $(\mathbf{k}_i - \underline{\mathbf{k}})^{ns}$, which, by definition, do not affect the calibration data set (and thus the precision of calibration). Due to model non-linearity, solution space and null-space can not be separated completely, resulting in some solution space component in $(\mathbf{k}_i - \underline{\mathbf{k}})^{ns}$, which alters the model-measurement fit of the parameter sets created in step 4

(usually to a less-calibrated fit). To alleviate for this, a final step is added to the process:

5. Re-calibration of the parameter sets $\underline{\mathbf{k}} + (\mathbf{k}_i - \underline{\mathbf{k}})^{ns}$ to generate $\underline{\mathbf{k}} + (\mathbf{k}_i - \underline{\mathbf{k}})^{ns \rightarrow rc}$.

Again, for a linear model, this re-calibration step would only affect the solution space of the model, thus keeping the null-space deviation created earlier fully intact. Dealing with model non-linearity, this can not be assured. Therefore, while re-calibration step is necessary to ensure that the random parameter sets generated calibrate the model sufficiently, this step can also affect the null-space deviation of the newly generated parameter sets. Too rigid a re-calibration could thus potentially falsely narrow the variability of the random parameter sets. Therefore, the recalibration has to be done cautiously, trying to generate sufficiently calibrating parameter sets while not narrowing the variability of the total of all sets too much. This process can therefore not capture the total variability of potential calibrating parameter sets due to model non-linearity, but only approach it. Nonetheless, the method is much less time-consuming that generating calibrated parameter sets from differing starting points by conventional calibration, especially in conjunction with surrogate models due to their low run times.

4.5.4 Robust data worth analysis

For the application of a robust data worth analysis method in this thesis, I combine the first-order second-moment uncertainty and data worth estimation techniques presented at the beginning of this section with the multiple calibrating parameter sets generated as presented above. I undertake this combination of first-order second-moment data worth estimation and efficient calibrating parameter set generation through the following steps:

1. Generation of a (base) parameter set $\underline{\mathbf{k}}$ by calibration of the model.

2. Generation of i random null-space perturbed parameter sets $\underline{\mathbf{k}}+(\mathbf{k}_i-\underline{\mathbf{k}})^{ns \rightarrow rc}$, which calibrate the model to a sufficient degree, via the method presented directly above.

3. Calculation of the basic predictive uncertainty $\sigma^2_{s_j|\mathbf{h}}(i)$ for j predictions of interest s_j for each of the i parameter sets $\underline{\mathbf{k}} + (\mathbf{k}_i - \underline{\mathbf{k}})^{ns \to rc}$.

4. Calculation of the worth of existing, "future" and "parametric" data as increase or decrease of the basic predictive uncertainties as explained above. This can be done for all (combinations of) parameters and measurements of interest.

5. Aggregation of the i different data worth results for each prediction of interest (and each data type) by calculation of mean and 90 % quantiles of data worth. Other statistical measures could obviously be applied here, as well.

The resulting data worth analysis, while based on first-order second-moment methods, creates more robust data worth estimations due to an approximation of the effects of parametric uncertainty while remaining reasonably time-consuming.

5 Results and discussion

In this chapter, I present the results of the application of the different methods presented above to the complex benchmark and three different surrogate models described earlier.

First, I start with the examination of the explicit boundary treatment in POD in section 5.1. In this section, I present the successful application of the method to eliminate boundary errors in different synthetic test models (see Gosses et al. (2018) in Appendix A) for Dirichlet (5.1.1) and Neumann (5.1.2) boundary conditions. Furthermore, I depict the reduction in errors at Cauchy boundaries through the method (5.1.3) and summarize the overall results for the explicit treatment of boundary conditions in POD with the new method (5.1.4).

Second, I present the analysis of simplification error for all three surrogate models in comparison to the complex model for a variety of model predictions in section 5.2. I discuss the simplification error estimates for the three surrogate models individually, starting with the simplified MODFLOW model $SM_{1,sMm}$(5.2.1), followed by the POD model $SM_{2,POD}$ (5.2.2) and the suite of ANNs $SM_{3,ANN}$ (5.2.3). I end the section by interpreting common features and differences regarding the simplification errors of the three different surrogate models in summary (5.2.4).

Third, I show the results of the data worth analysis with two surrogate models and the complex model in comparison in section 5.3. I group the results for existing (5.3.1), "future" (5.3.2) and "parametric" (5.3.3) data individually and end with a summary of the results and interpretation of the data worth analysis (5.3.4).

As a final section 5.4, I repeat the summary interpretation of the results for the three individual objectives of the thesis and discuss the overall findings of this thesis in regards to surrogate modeling of different types and its application for uncertainty and data worth analysis in groundwater.

5.1 Explicit treatment of boundary conditions in POD

As summarized in section 4.2 and presented in detail in Gosses et al. (2018), I developed an extension to POD for the explicit treatment of boundary conditions (eb-POD) on the basis of Gunzburger et al. (2007) due to inaccuracies in boundary representation in classical POD applied in hydrogeology. Based on an explicit representation of Dirichlet boundary conditions in POD, I extended the method to the other two boundary condition types in groundwater modeling, Neumann and Cauchy boundaries. To test the methodology for all three boundary types and demonstrate its performance in comparison to both complex model and classic POD, I developed several synthetic test models of increasing (boundary-related) complexity, mostly in 1D. While I present the results below, the precise setup of the test models can be found in Gosses et al. (2018) in Appendix A.

5.1.1 (Variable) Dirichlet boundaries

As the original method developed by Gunzburger et al. (2007) extends POD for Dirichlet boundary conditions, I start by applying the eb-POD method to Dirichlet boundaries in groundwater POD in this subsection. As stated in Gosses et al. (2018), "Dirichlet boundaries can potentially be handled correctly by standard POD with several restrictions". Nonetheless, these restrictions are often cumbersome or time-consuming to implement. Figure 3 shows the maximum groundwater head error over time (in comparison to the complex model) of the eb-POD model at the boundary for a simple 1D test case with a variable groundwater head (i.e. Dirichlet) boundary. As is apparent, variable Dirichlet boundaries can be reproduced by eb-POD with errors far below any significance demand. Table 3 shows the maximum error and NRMSE of the eb-POD model in the general reproduction of the groundwater head in the model area, which is about 1.5 cm. Note that the test model has a horizontal extend of 10 km and a vertical groundwater gradient of over 3 m. Therefore, the divergence between complex model and eb-POD model is quite small in relative terms. Due to the difficulties of representing variable Dirichlet boundaries with POD, I built no classic POD model for comparison in this case.

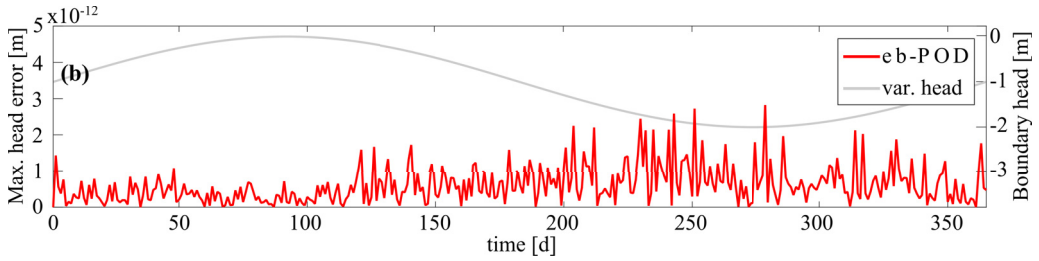

Figure 3: eb-POD for one variable Dirichlet boundary (groundwater head). Absolute maximum groundwater head error at the boundary over time for eb-POD and change in head at the variable groundwater head boundary. Figure modified from Gosses et al. (2018).

5.1.2 Neumann boundaries

Next is an application of the eb-POD method to a synthetic 1D model with several Neumann boundaries. Neumann boundaries signify specified fluxes and are used to represent well pumping and infiltration from unconnected rivers, for example. As discussed in subsection 4.2.3, eb-POD should handle Neumann boundaries without error. Figure 4 shows several different errors over time (again, in comparison to the complex model) for both classic POD and the eb-POD model. Figure 4a shows the maximum groundwater head error over time, which is also summarized in Table 3. It is apparent that the eb-POD model is less accurate than the classic POD model in regard of the overall reproduction of the groundwater head field. This is due to the decrease in size of the actual projection matrix in eb-POD due to the boundary representation, as explained in subsection 4.2.4. Looking at the relative flux errors through the boundary conditions (Figure 4b&c show two different pumping wells, while Figure 4d shows the river–groundwater exchange flux error) demonstrates the gained accuracy at the boundaries for eb-POD, though. The classic POD model results in measurable relative flux errors of over 5 % at the river boundary, and even some correlation of errors between the different boundaries. The explicit treatment of the Neumann boundaries through eb-POD totally eliminates these errors, successfully proving the extension to the methodology to represent Neumann boundaries error-free.

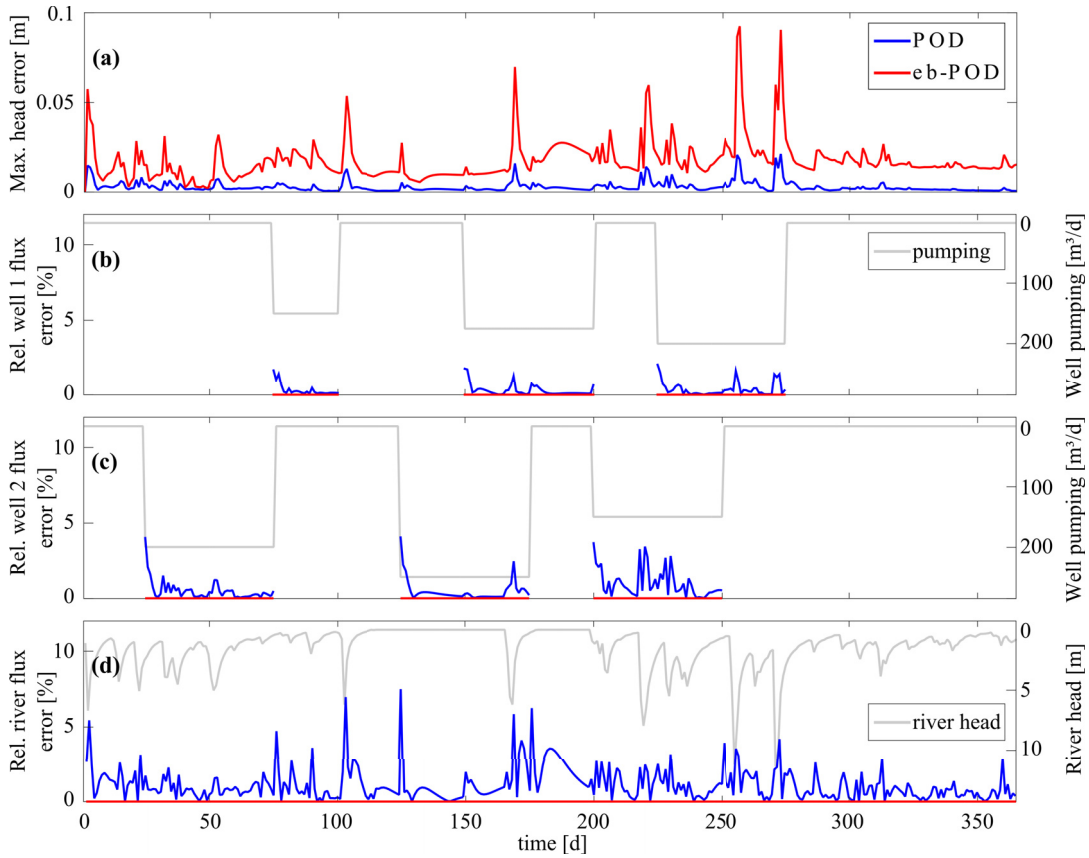

Figure 4: POD and eb-POD for several Neumann boundaries (two wells, one river). (a) Absolute maximum groundwater head error over time for both POD and eb-POD surrogates. (b), (c) Relative well flux errors over time along with the appropriate pumping rates at both pumping wells. (d) Relative river flux error over time and river head. Figure modified from Gosses et al. (2018).

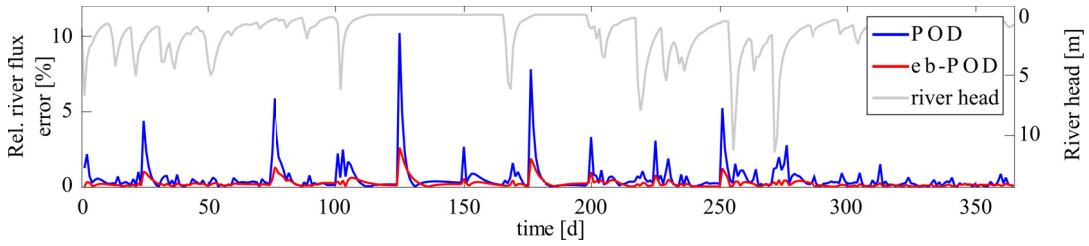

Figure 5: POD and eb-POD for one Cauchy boundary: relative flux error over time, along with the river head. Figure taken from Gosses et al. (2018).

5.1.3 Cauchy boundaries

The final type of boundary condition are non-linear Cauchy boundaries. The model setup to test the eb-POD method for Cauchy boundaries is similar to the one directly above, but with a lower river bed directly connecting the river to the groundwater. This changes the boundary to a Cauchy boundary, which is non-linear and can be represented by a groundwater-head-dependent and groundwater-head-independent part (see subsection 4.2.3 and Gosses et al. (2018)). The groundwater-head-independent part is exactly like a Neumann boundary in its representation in the model, though, allowing the treatment through eb-POD. Figure 5 shows the relative river flux error over time for both POD and eb-POD. In comparison to Figure 4d, the overall errors are bigger, and the eb-POD error is not 0, due to the inability to explicitly treat the groundwater-head-dependent part of the Cauchy boundary. Nonetheless, the eb-POD method greatly reduces the relative river flux error through the Cauchy boundary in comparison to the classic POD method, thus proving the potential of applying eb-POD to Cauchy boundaries. Table 3 shows that this error reduction in the boundary flux again comes at a cost of slightly higher overall groundwater head field errors.

5.1.4 Applying eb-POD: summary

The simple test cases presented above prove the successful application of explicit treatment of all three boundary condition types in groundwater modeling by my newly developed eb-POD method. Gosses et al. (2018) includes an application

Table 3: General groundwater head errors for POD and eb-POD models compared to the complex models (all values in [m]). Table modified from Gosses et al. (2018).

ROM	**Dirichlet**		**Neumann**		**Cauchy**	
	Max. error	NRMSE	Max. error	NRMSE	Max. error	NRMSE
POD	–	–	0.0215	0.0010	0.0178	0.0006
eb-POD	0.0148	0.0048	0.0904	0.0051	0.0596	0.0034

to a more complex, 3D test case as well, demonstrating the portability of the method to more complex cases. Classic POD can potentially represent Dirichlet boundaries error-free, but the application is cumbersome. Neumann and Cauchy boundary fluxes are often highly erroneous in classic POD, as shown above. eb-POD is successful in eliminating errors at Dirichlet and Neumann boundaries, and reducing Cauchy boundary errors by their groundwater-head-independent part. Due to a reduction in projection matrix size necessary for this boundary treatment, general groundwater head errors are slightly higher for eb-POD in comparison to classic POD. As eb-POD can be easily applied and the trade-off in general accuracy is quantifiable, it gives the modeler the chance for purposeful decision in location and type of these errors.

In summary, the major results and interpretations of testing the eb-POD method for all three types of boundary conditions are:

- eb-POD is successful at eliminating errors of classic POD at Dirichlet and Neumann boundaries.
- Errors of classic POD at Cauchy boundaries are reduced through eb-POD by their groundwater-head-independent part.
- General groundwater head errors of eb-POD are slightly higher than of classic POD due to the reduction in projection matrix size for eb-POD. This can be easily quantified, though.
- Easy application and quantifiability of gains in boundary representation and trade-off in general head reproduction with eb-POD allow purposeful decision-making for its application.

5.2 Quantifying model simplification error

After successfully proving the advantages of eb-POD in representing boundary conditions in POD surrogates, I built three different surrogate models, based on a complex benchmark model, as presented in section 4.1. To gauge the performance of all three surrogate models as predictors in comparison to the complex model, I quantified model simplification error for different predictions following the methodology presented in sections 4.3 and 4.4. The evaluation of the model simplification error for different surrogate models in comparison to a complex groundwater model for a real-world test case, as presented below, was published in Gosses and Wöhling (2019). I used the paired surrogate–complex model approach to estimate prediction-specific simplification errors first presented by Doherty and Christensen (2011) for three surrogate models: a simplified MODFLOW model ($SM_{1,sMm}$, 5.2.1), a proper orthogonal decomposition model ($SM_{2,POD}$, 5.2.2) and a suite of artificial neural networks ($SM_{3,ANN}$, 5.2.3). For all three surrogates, I extended the method by Doherty and Christensen (2011) for the use with a calibrated complex model to facilitate analysis of (potential) simplification error underestimation and gain some insights into the prediction-specific complex model errors as shown in subsection 4.4.4. Furthermore, I applied the comparison method via parameter composition developed by Watson et al. (2013) to analyze the extend of parameter surrogacy for the $SM_{1,sMm}$ model (5.2.1). Finally, I draw some overall conclusions regarding the different surrogate modeling categories and their predictive performance (5.2.4).

5.2.1 Simplified MODFLOW model: $SM_{1,sMm}$

I start with the results of the simplified MODFLOW model $SM_{1,sMm}$ presented in subsection 4.1.3. The $SM_{1,sMm}$ is a spatially and parametrically simplified version of the complex benchmark model also using MODFLOW. I created scatter plots of simplification error for a multitude of different predictions from the three main groups of minimum groundwater head predictions, mean river–groundwater exchange flux predictions and mean spring flux predictions. The full results for the $SM_{1,sMm}$ model can be found in Gosses and Wöhling (2019) in Appendix A.

In general, the results can be sorted into two main groups, along with two minor ones. I present and interpret abbreviated results according to this grouping here. Figure 6 shows an exemplary scatter plot for one prediction belonging to each of these groups.

Figure 6a, labeled "calibrated" prediction, shows the scatter plot of CM vs. $SM_{1,sMm}$ model predictions for the mean Spring Creek flux. For this spring, discontinuous measurements were available and used as part of the calibration data set for the models. As Figure 6a shows, this lead to small simplification errors (indicated by the vertical extend of the error bars) for the $SM_{1,sMm}$ model for the accompanying model prediction, which is verified by the $[\underline{s}_{\mathbf{o(h)}}; s_{\mathbf{h}}]$ pair lying inside the error band. Thus, simplification errors are small for predictions in locations where measurement data was available and used as part of the calibration (i.e. that are known very well to the model). This conforms to general expert assessment, as the surrogate model is forced to reproduce data of the calibration data set well through the calibration process, which should counteract possible deterioration of its ability to make such "calibrated" predictions through the simplification process.

The second major group, labeled "unknown" predictions, is exemplarily shown in Figure 6b for the prediction of minimum groundwater head at the P Neal well. For this well, no data was used in the calibration process. As a result, simplification error is larger than for predictions of the first group. The more different predictions are to the calibration data set in type and location, the larger the simplification errors for the $SM_{1,sMm}$ model, in general. Nonetheless, the $[\underline{s}_{\mathbf{o(h)}}; s_{\mathbf{h}}]$ pair is usually inside the error band, signifying that simplification errors are not underestimated. Again, it is reasonable that such "unknown" predictions are more strongly affected by model simplification, thus exhibiting higher simplification error, than the previously presented "calibrated" predictions.

For certain predictions, other interesting information can be gained from the scatter plots. Figure 6c shows the scatter plot for the prediction of minimum groundwater head in the Wratts Rd. well, labeled an "eastern" prediction. While the small simplification error is in agreement with the prediction (Wratts Rd.

groundwater head data is used in the calibration data set), the $[\underline{s}_{\mathbf{o(h)}}; s_{\mathbf{h}}]$ pair lying outside the error bands suggests an underestimation of simplification error. This groundwater well is located in the eastern part of the aquifer, where the aquitard is appearing on top of the aquifer, forcing the groundwater up and into the springs (cf. Figure 1). I conclude this to be an area in need of high spatial (and parametrical) variability to satisfactorily resolve these effects. As this variability is missing from the $SM_{1,sMm}$ model, it is not able to reproduce the corresponding predictions with certainty.

Figure 6d, showing the mean river–groundwater exchange flux prediction between gauging stations SH6 and Wratts Rd., is an example of another type of conspicuity. For this prediction, the slope of the regression line is below 0.5, showing that the $SM_{1,sMm}$ model is generating more variable results than the CM model based solely on its precalibration uncertainty. The high extend of horizontal scatter suggests effects of parameter surrogacy (thus used as a label), which I introduced in section 4.3. Parameter surrogacy means that the $SM_{1,sMm}$ model parameters are stand-ins for certain processes and parameters of the complex model that are different to their respective intended purpose. This could be hydraulic conductivity zones in the surrogate model also taking on the function of boundary conductances, for example. I tested this by computing the composition of $SM_{1,sMm}$ model parameters in regards to the CM model parameters by Eq. (17). This is shown in Figure 7, with CM model parameters grouped in conjunction with $SM_{1,sMm}$ model parameters for ease of analysis. One can detect that no $SM_{1,sMm}$ model parameter is mainly contributed to by its complex model counterpart. Instead, for most $SM_{1,sMm}$ model parameters, a wide range of complex model parameters contribute, belonging to various parameter groups, from hydraulic process representation to boundary conditions. This is a strong sign of parameter surrogacy in the $SM_{1,sMm}$ model, where the parameters thought to be representing certain real-world processes (and such being limited in their respective parameter ranges by expert knowledge) are instead composites of a variety of these processes. In conclusion, the $SM_{1,sMm}$ model parameters should be treated as "effective" parameters, allowed to vary freely in the calibration, instead of being

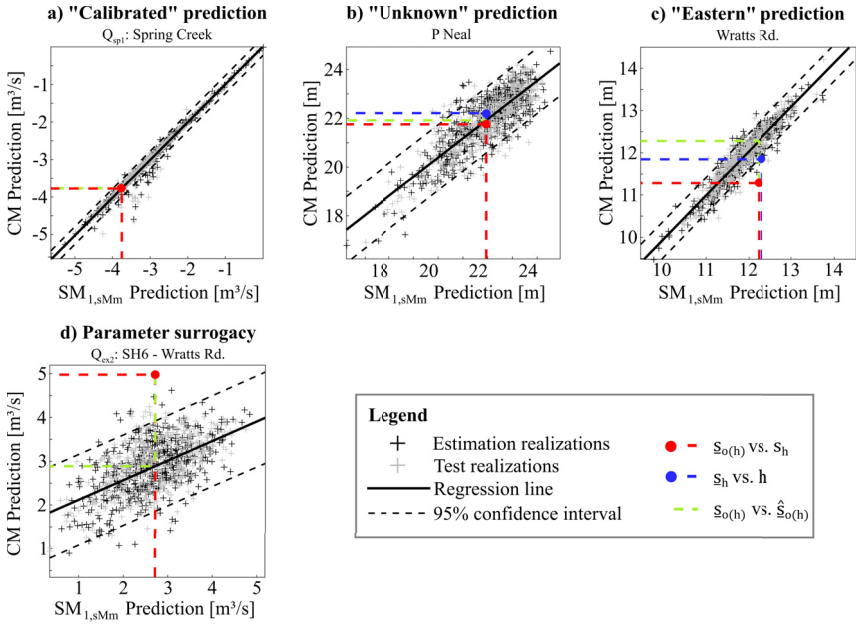

Figure 6: Scatter plots of CM vs. $SM_{1,sMm}$ model predictions. a) mean Spring Creek flux prediction at type and location of measurements pertaining to the calibration data set. b) minimum groundwater head at validation well P Neal. c) minimum groundwater head at Wratts Rd. well lying in the eastern part of the model problematic for the $SM_{1,sMm}$ model. d) mean river–groundwater exchange flux prediction between gauging stations SH6 and Wratts Rd. as an example of parameter surrogacy influence. Figure modified from Gosses and Wöhling (2019).

tied to real-world processes by range. I infer that this has its strongest effects for predictions differing widely in location and type from the calibration data set and directly dependent on certain model parameters, as is the case for the prediction of Q_{ex2}.

5.2.2 POD surrogate model: $SM_{2,POD}$

Next is the POD surrogate model, $SM_{2,POD}$, presented in subsection 4.1.4. $SM_{2,POD}$ is a highly reduced version of the linearized complex model, solving the equation system on a subspace of size 7 instead of 6,360. Due to the POD method, it retains the original parameterization, though. As with the $SM_{1,sMm}$ model results, the scatter plots for all different predictions for the $SM_{2,POD}$ model can be

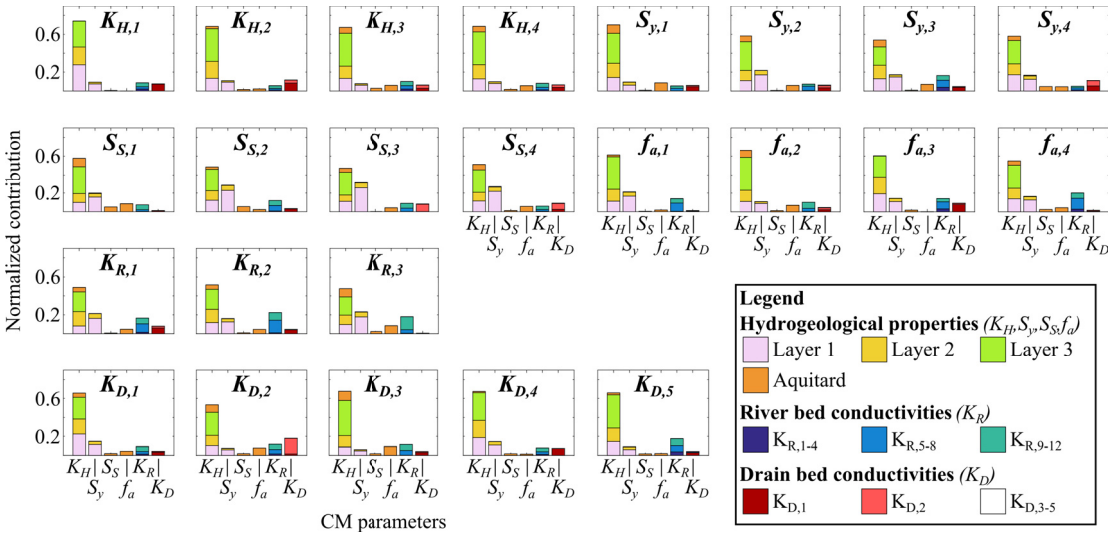

Figure 7: Normalized contribution of the (grouped) CM model parameters to the 24 $SM_{1,sMm}$ model parameters. Figure taken from Gosses and Wöhling (2019).

consolidated to certain groups, which are again exemplarily depicted in Figure 8.

The first group, shown in Figure 8a, are again the "calibrated" predictions. Similar to the $SM_{1,sMm}$ model, the $SM_{2,POD}$ model shows smaller simplification errors for predictions pertaining to the calibration data set in type and location. Figure 8b shows that "unknown" predictions, i.e. predictions different in type and location to the calibration data set, demonstrate larger simplification errors again, as for the $SM_{1,sMm}$ model.

Looking at the minimum groundwater head prediction at Wratts Rd. in Figure 8c, we can see differences between the $SM_{2,POD}$ model and $SM_{1,sMm}$ model. While again labeled an "eastern" prediction, the simplification error is not underestimated, as apparent from the $[\underline{s}_{\mathbf{o(h)}}; s_{\mathbf{h}}]$ pair inside the error band. Nonetheless, the scatter shows a curious pattern, with higher $SM_{2,POD}$ model prediction values clustered vertically, indicating a disconnect between CM and $SM_{2,POD}$ model prediction. I concur that this effect stems from the linearization applied to the model necessary for the application of POD, which has its strongest influence in the eastern part of the model, where it affects the thinning phreatic aquifer, its unconfined groundwater table and the interaction with the spring cells.

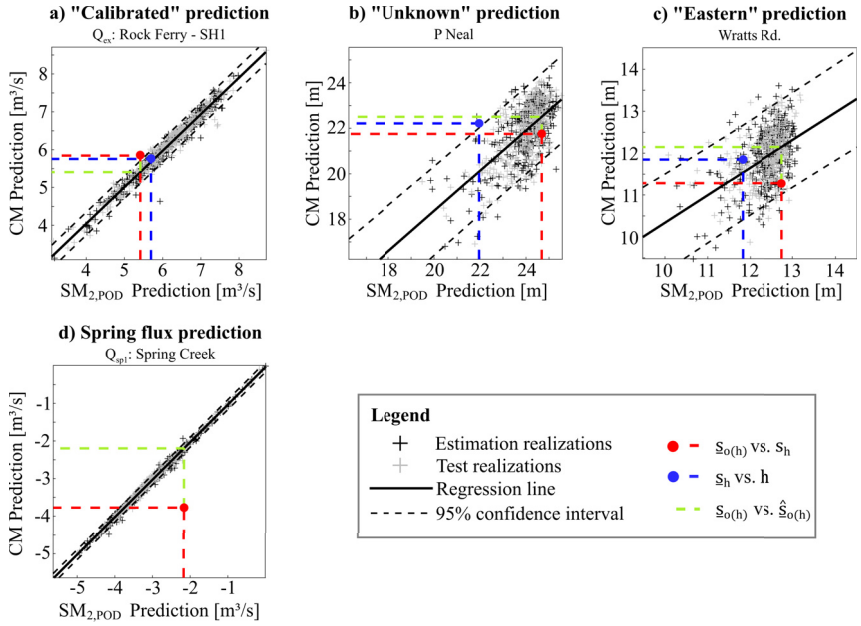

Figure 8: Scatter plots of CM vs. $SM_{2,POD}$ model predictions. a) mean total river–groundwater exchange flux, used in the calibration data set. b) minimum groundwater head at validation well P Neal. c) minimum groundwater head at Wratts Rd. well lying in the eastern part of the model problematic for POD. d) mean Spring Creek flux prediction, showing the error underestimation for spring flows by the $SM_{2,POD}$ model.

The linearization effects on spring fluxes can further be seen in 8d. While the simplification error for the mean Spring Creek flux prediction is shown to be very small, which is in line with this prediction being a "calibrated" one, the $[\underline{s}_{\mathbf{o(h)}}; s_{\mathbf{h}}]$ pair's location far outside the error band suggests a strong underestimation of the simplification error. As this is consistent for all spring flux predictions, I assume this comes from the linearization as well, limiting the range of spring fluxes in $SM_{2,POD}$ model realizations.

5.2.3 ANN surrogate model: $SM_{3,ANN}$

The last type of surrogate models is the suite of ANN models, called $SM_{3,ANN}$. These differ dramatically from the two previous surrogate models, as a single

ANN model was built and calibrated for each data series of interest, as explained in subsection 4.1.5. Thus, $SM_{3,ANN}$ shows very different simplification error scatter plots than the first two surrogate models. Figure 9a–d show scatter plots for four different predictions: a minimum groundwater head in a well belonging to the calibration data set (a), a minimum groundwater head at a validation well (b), a spring flux prediction (c) and a river–groundwater exchange flux prediction (d).

For all four predictions, disregarding of type or location, simplification errors for the $SM_{3,ANN}$ models are very small. Examination of the $[\underline{s}_{\mathbf{o(h)}}; s_{\mathbf{h}}]$ pairs suggests that these small simplification errors are correct, as the pairs lie within or very close to the uncertainty bands. This discrepancy to the other two surrogate models can be explained by the difference in setup between them and the $SM_{3,ANN}$ models. Where the $SM_{1,sMm}$ model and the $SM_{2,POD}$ model are distributed, single models for all predictions, an individual $SM_{3,ANN}$ model was set up and calibrated for each data type and location corresponding to a prediction of interest. This gives the $SM_{3,ANN}$ models the advantages of a) only having to simulate this one time series without having to incorporate trade-offs to other, potentially adverse measurements, and b) always having the data corresponding to its prediction available in the calibration data set. As ANNs need sufficient training data corresponding to the prediction of interest, this setup was necessary. This was possible in the scope of this study, as the "measurement data" used for calibration of the surrogates are indeed complex model simulations, which are readily available everywhere. This is obviously not the case in real-world application, thus severely limiting the advantages of the $SM_{3,ANN}$ models.

5.2.4 Surrogate comparison: simplification errors in model predictions

Comparing the simplification errors for the three surrogate models along the wide range of model predictions allows for some classification of and general statements about the surrogate models and their respective types.

The $SM_{3,ANN}$ models stand out from the other two surrogate models by the very small simplification errors for all types of predictions. As explained above, this is due to the setup of the $SM_{3,ANN}$ models, where a single surrogate was

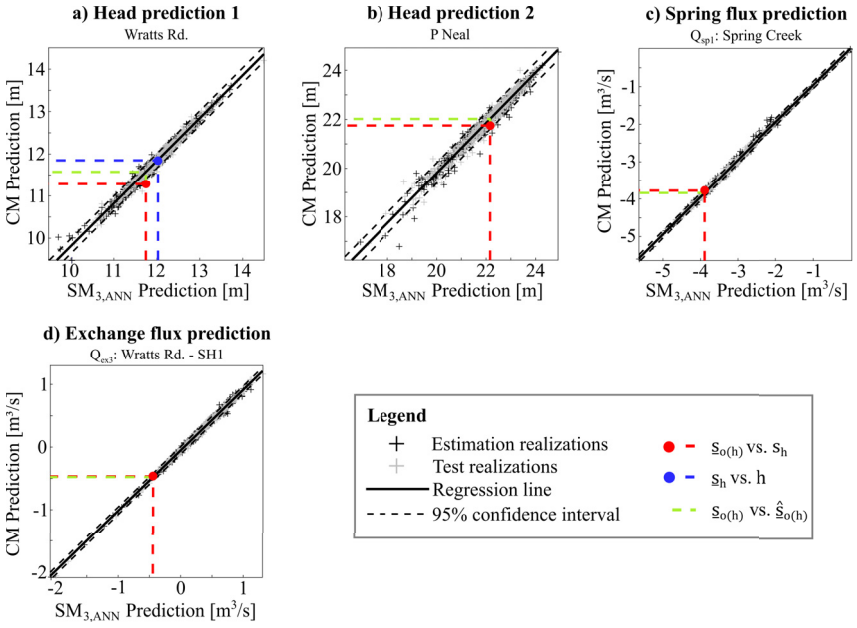

Figure 9: Scatter plots of CM vs. $SM_{3,ANN}$ model predictions. a) minimum groundwater head at (calibration data set) Wratts Rd. well. b) minimum groundwater head at validation well P Neal. c) mean Spring Creek flux prediction. d) mean river–groundwater exchange flux prediction between gauging stations Wratts Rd. and SH1. Figure modified from Gosses and Wöhling (2019).

built and trained for each data time series. This limits the possible applications for ANNs in groundwater prediction, but in cases where data is available, they are clearly performing very well in reproducing CM model outputs without much error.

Both the $SM_{1,sMm}$ model and the $SM_{2,POD}$ model show similar, small, simplification errors for predictions pertaining to the calibration data set. In general, they both show an increase in simplification error the farther a prediction is in type and location to the calibration data set. Furthermore, both show difficulty in reproducing certain model predictions in the eastern model domain either accurately ($SM_{2,POD}$) or believably ($SM_{1,sMm}$).Where the high simplification errors of the $SM_{2,POD}$ model in the eastern part of the model likely stem from the underlying linearization, the probable underestimation of simplification error in this model part by the $SM_{1,sMm}$ model is due to its missing variability in spatial and parametrical resolution in this highly heterogeneous area. The $SM_{1,sMm}$ model also shows some effects of parameter surrogacy for certain predictions, where the limited parameterization scheme of the surrogate model leads to mixing contributions of CM model parameters onto their surrogate counterparts, thus limiting their ability to represent real-world processes. While the analysis produced small simplification errors for spring flux predictions with the $SM_{2,POD}$ model, the $[\underline{s}_{\mathbf{o(h)}}; s_{\mathbf{h}}]$ pairs suggest some error underestimation here, likely due to the strong linearization effects on these boundary fluxes. Regarding run time reduction, the $SM_{1,sMm}$ model runs about 5–10 times faster than the CM model, the $SM_{2,POD}$ model is about 30 times faster, and the $SM_{3,ANN}$ models run about 1,000 times faster.

In summary, the following statements about the three surrogate models and their predictive performance can be made:

- ANNs can reproduce predictions pertaining to existing measurement data sets used in training with high accuracy. They are not generally suitable for predictions without such measurements, though.

- The simplified MODFLOW surrogate is restricted in its ability to correctly reproduce complex model predictions by the type of simplification applied.

Generally, simplification errors are higher the more predictions differ in type and location to the calibration data set. Structural simplification can have adverse effects on surrogate predictions in areas in need of high spatial resolution. Parametric simplification has similar effects and can furthermore lead to parameter surrogacy, which decouples surrogate model parameters from real-world equivalency.

- The POD surrogate is similar to the simplified MODFLOW model in certain aspects, as simplification errors are higher the more predictions differ in type and location to the calibration data set for them, too. Contrary, it does not suffer from parameter surrogacy, as the POD surrogate retains the complex model parameterization. Necessary model non-linearity has adverse effects on POD predictions due to the linearity of the method.

5.3 Robust data worth analysis using surrogate models

After presenting progress on the methodology of POD in section 5.1 and analysis on the cost of model simplification regarding predictive behavior of the surrogate models in section 5.2, I come to the results of the data worth analysis using the surrogate models. As depicted in section 4.5, I employed first-order second-moment uncertainty estimation methods developed by Doherty (2016) to estimate data worth in regard to increasing or decreasing postcalibration uncertainty estimates. As these methods depend on the calibrated parameter set, I employed null-space techniques (Doherty, 2016) to generate a set of calibrated parameter set for a combined, more robust data worth estimate taking into account parametric uncertainty. In the following, I present results of the application of this robust data worth analysis for CM, $SM_{2,POD}$ and $SM_{1,sMm}$ models regarding several model predictions. I could not use the $SM_{3,ANN}$ models for data worth analysis, as they do not allow for the implementation of "future" data and do not have physical parameters either, thus severely limiting their possible applications for data worth analysis. I analyzed the data worth regarding existing measurements used in the calibration data set (5.3.1), regarding "future" groundwater head measurements

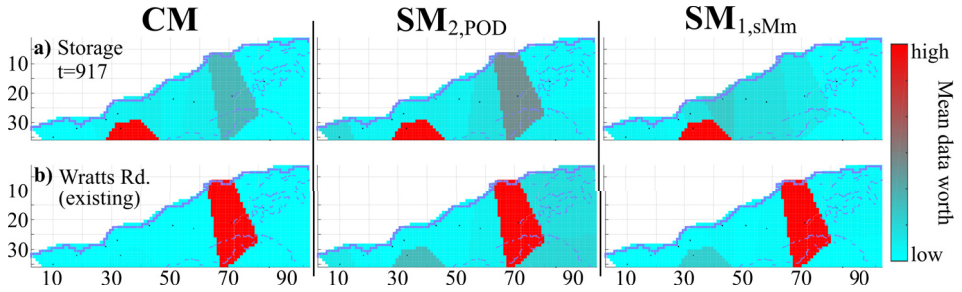

Figure 10: Mean worth of existing data (shown as zones) for CM, $SM_{2,POD}$ and $SM_{1,sMm}$ models regarding predictions of: a) cumulative aquifer storage at t=917 days, and b) minimum groundwater head at Wratts Rd. well.

(5.3.2) and regarding "parametric" data as (potential) knowledge of model parameters (5.3.3). Again, I end with summarizing the implications of the results regarding the performance of the two surrogate modeling categories employed for data worth analysis under consideration of parametric uncertainty (5.3.4).

5.3.1 Worth of existing data

I start by presenting results of the analysis of both surrogate model ($SM_{1,sMm}$ and $SM_{2,POD}$) and the complex model regarding the worth of the existing groundwater head measurement data used in the calibration data set. Figure 10 shows the mean data worth for all three models for two different predictions. Data worth is mapped as the mean value of all 100 calculations (based on the different parameterizations) and plotted in zones around the nearest groundwater well of the calibration data set – note that no absolute values are given, as these are a) highly uncertain and b) do not scale appropriately for the three different models.

For both the prediction of aquifer storage at t=917 days and the mean groundwater head at the Wratts Rd. well (for which data is used in the calibration data set), $SM_{2,POD}$ and $SM_{1,sMm}$ models are able to reproduce the zone of high data worth that the CM model identifies. This suggests that both surrogate models are capable of evaluating the existing monitoring network, even for predictions with high simplification errors associated, as presented earlier.

5.3.2 Worth of "future" data

After assessment of the currently available groundwater head data, I extended the analysis for "future", not-yet-existing groundwater head data. Figure 11 shows the mean worth of "future" groundwater head measurements along a grid (represented as dots in the graph) for six different predictions.

While there is some divergence between CM and $SM_{2,POD}$ models, the overall areas of high data worth are generally identified quite well by the $SM_{2,POD}$ model. Thus, the POD surrogate is able to reproduce worth of "future" data in comparison to the CM results, disregarding its initial simplification error for the prediction of interest.

There are some overlaps between the areas of high data worth identified by $SM_{1,sMm}$ and CM models for the storage (a) and groundwater head (b,c) predictions, as well as for the spring flow (f). The $SM_{1,sMm}$ model fails to properly reproduce the areas of high data worth found by the CM model for river–groundwater exchange fluxes (d,e) , though, and identifies a major area of importance for the spring flow (f) that can not be found in the CM model map. This suggests that worth of "future" data generated with the $SM_{1,sMm}$ surrogate is not generally trustworthy. As the idea behind using a surrogate for data worth analysis would be to omit the computations with the complex model, I deem the $SM_{1,sMm}$ model unable for this task in regard to "future" data.

Figure 12 shows the same map for two of the model predictions, but plots the 90 % quantiles of data worth (calculated from the 100 different parameterizations). This gives some indication on the necessity of the robust method to take parametric uncertainty into account. For the spring flow prediction (a), the area of high data worth variance coincides with the area of high data worth and is similar between the CM model and the surrogates. This suggests that for this prediction, a fewer amount of data worth realizations (or even a single one with the original calibrated parameter set) would have been enough to identify the area of high data worth for all three models, making the additional runs obsolete. Thus, parametric uncertainty seems to have no effect on the worth of "future" groundwater head data for this specific prediction.

Figure 12b shows a different case: while the CM model again generally identifies the area of high data worth in the mapping of 90 % data worth quantile (suggesting that much fewer runs would suffice), the surrogate models do not. Despite identifying areas of high data worth variances at completely different places, both $SM_{2,POD}$ and $SM_{1,sMm}$ models were able to correctly identify the area of high data worth correctly for this prediction (cf. Figure 11b). This shows the necessity of the robust method to account for parametric uncertainty when using surrogate models for data worth analysis for such predictions. I identified this for several predictions, especially for the $SM_{2,POD}$ model (results not shown), suggesting that while I deem the $SM_{2,POD}$ model a suitable surrogate for data worth analysis for "future" data, this only holds true if taking parametric uncertainty into account via the robust data worth method.

5.3.3 Worth of "parametric" data

Finally, I analyze the ability of the surrogate models to reproduce the data worth findings of the CM model in regard to parameter knowledge, i.e. "parametric" data. Figure 13 shows the results for three different predictions, only mapping the most influential parameters of each model. It is important to keep in mind here that the $SM_{2,POD}$ model keeps the full parameter set of the CM model, while the $SM_{1,sMm}$ model uses aggregates of the spatially distributed parameter fields of hydraulic conductivity, specific yield and river bed conductivity as presented in section 4.1 in Table 2.

For the prediction of aquifer storage in Figure 13a, the CM model identifies some area of specific yield in layer 1 as influential parameters, which makes sense in light of those values being the key factor in determining aquifer storage. The $SM_{2,POD}$ model successfully identifies the same parameter group, and is able to generally identify the area, as well. The $SM_{1,sMm}$ model does not have spatially distributed parameter fields, but it still has a parameter that encapsulates the same process: $S_{y,1}^{sMm}$, the specific yield of layer 1 in the $SM_{1,sMm}$ model. Figure 13a shows that the data worth analysis identifies $S_{y,2}^{sMm}$, the specific yield of layer 2, as the (only) influential parameter for this prediction, though.

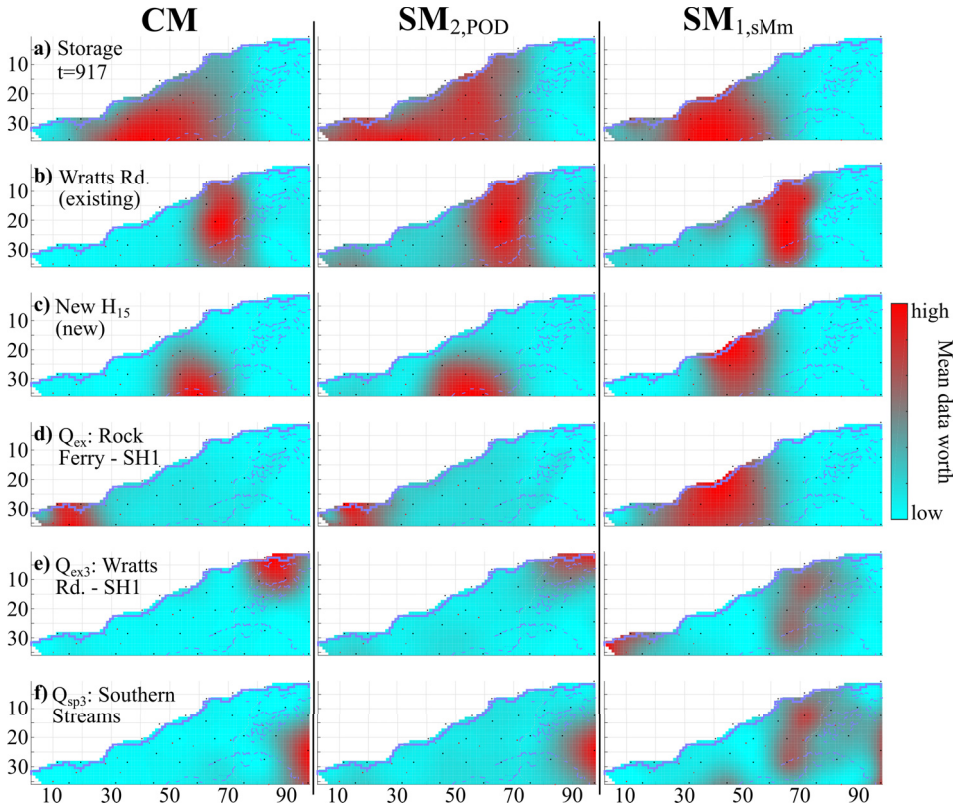

Figure 11: Mean worth of "future" data (black dots) for CM, $SM_{2,POD}$ and $SM_{1,sMm}$ models regarding predictions of: a) cumulative aquifer storage at t=917 days, b) minimum groundwater head at Wratts Rd. well, c) minimum groundwater head at potential new well 15, d) mean total river–groundwater exchange flux, e) mean river–groundwater exchange flux between gauging stations Wratts Rd. and SH1, and f) mean Southern Streams spring flux.

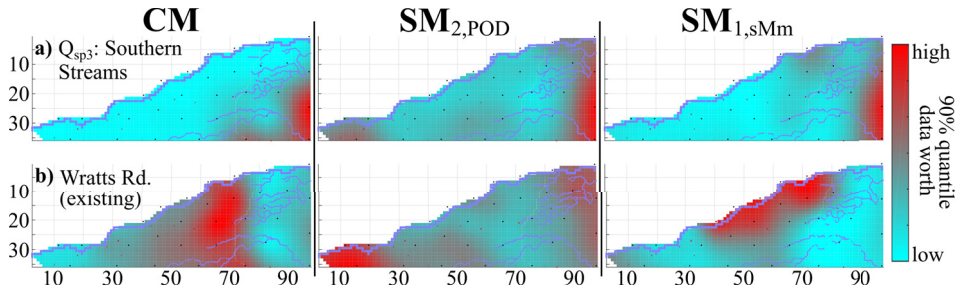

Figure 12: 90 % quantiles of worth of "future" data (black dots) for CM, $SM_{2,POD}$ and $SM_{1,sMm}$ models regarding predictions of: a) mean Southern Streams spring flux, and b) minimum groundwater head at Wratts Rd. well.

Figure 13b&c show the results for two different predictions. Again, the CM model identifies parameters as most influential that make sense from a hydrogeological standpoint (the river bed conductivity in the corresponding zone for a river–groundwater exchange flux (b) and the drain-bed conductivity of the corresponding spring for spring flux (c)), and the $SM_{2,POD}$ model is able to identify the same parameters. The $SM_{1,sMm}$ model does not only fail to identify the same parameters as influential as the CM model, but suggests parameters that do not logically make sense.

Going back to the findings about the high level of parameter surrogacy for the $SM_{1,sMm}$ model presented in section 5.2, it makes sense that the $SM_{1,sMm}$ model is not able to reproduce data worth of parameters. I tried to use the parameter contribution data generated in section 5.2 for a translation of the $SM_{1,sMm}$ model parameter data worth results, but this proved not successful (not shown here). The $SM_{2,POD}$ model showed to be a suitable surrogate for the analysis of worth of "parametric" data, though. I assume this is due to the $SM_{2,POD}$ model's retaining of original CM model parameters. Again, the $SM_{2,POD}$ model data worth results were accompanied by higher variance, though, suggesting the necessity of the robust method incorporating parametric uncertainty in successfully employing the $SM_{2,POD}$ model as a surrogate.

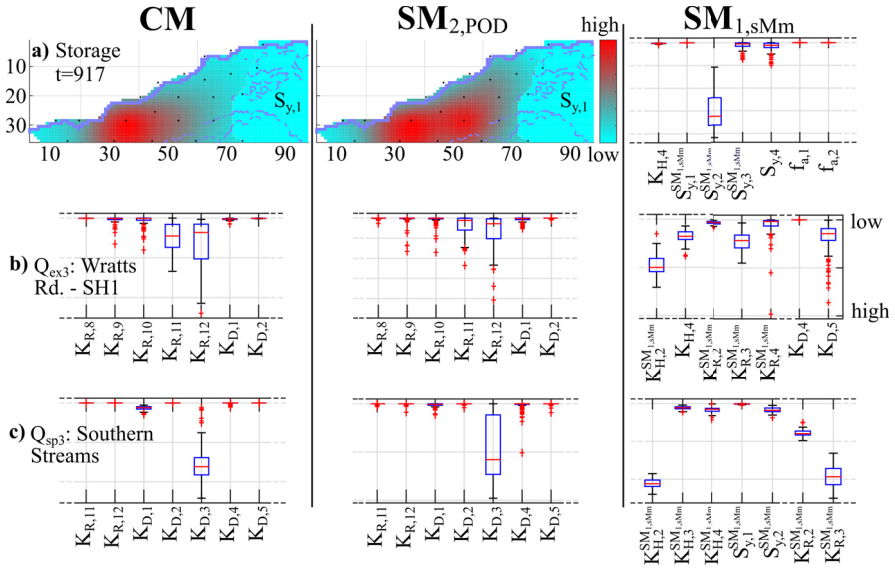

Figure 13: Mean worth of "parametric" data (only most sensitive shown) for CM, $SM_{2,POD}$ and $SM_{1,sMm}$ models regarding predictions of: a) cumulative aquifer storage at t=917 days, b) mean river–groundwater exchange flux between gauging stations Wratts Rd. and SH1, and c) mean Southern Streams spring flux.

5.3.4 Data worth with surrogate models: summary

Using surrogate models for data worth analysis instead of the complex model imposes some requirements on the surrogate models in question: they need to incorporate the full calibration data set (regarding the worth of existing data), must be physically distributed (regarding the worth of "future" data) and have somewhat physically-based parameters (regarding the worth of "parametric" data). This excludes the $SM_{3,ANN}$ models from undertaking predictive data worth analysis following the methodology presented in this thesis.

The $SM_{1,sMm}$ model was able to identify the most important existing measurement data for different predictions, suggesting it can be used as a surrogate for existing network analysis. Its difficulties to properly reproduce the worth of "future" data and the complete inability to identify important parameters disregarding of prediction of interest severely limit its applications for data worth analysis, though. In contrast, the $SM_{2,POD}$ model was able to reproduce all the

CM model findings in regard to data worth, independent of application type (existing, "future" or "parametric" data) and model prediction. This was even true for model predictions associated with high simplification errors or (presumably) underestimated simplification errors. The high variability associated with many of the data worth estimations by the $SM_{2,POD}$ model, often in areas of generally low data worth as well, suggests that this potential use of the $SM_{2,POD}$ model as a data worth surrogate is conditional on a robust method taking into account parametric uncertainty, though, as presented here.

In summary, results of applying first-order second-moment data worth analysis with several parameterization using surrogate models are:

- ANNs of chosen architecture are unsuitable for data worth analysis as they lack physical distribution of output and physically-based parameters.
- The simplified MODFLOW model reproduced worth of existing data correctly, but is limited in its use for "future" data worth analysis due to its structural simplification. The parametric simplification render it completely useless for the assessment of "parametric" data worth.
- The POD surrogate successfully reproduced worth of existing, "future" and "parametric" data as identified by the complex model. The high variance associated with many of its results suggest that this achievement is conditional on the application of a robust data worth method which takes parametric uncertainty into account.

5.4 Discussion

In this section, I quickly summarize the results of the different work done in this thesis again and discuss the major findings.

In regard to advancing POD surrogates in groundwater modeling, the presented eb-POD method successfully eliminates boundary errors at Dirichlet and Neumann boundaries and reduces errors at Cauchy boundaries, compared to basic POD. As shown in the results and discussed in the methodology, this comes at

the cost of a small increase in overall inaccuracy of the groundwater head representation of the eb-POD surrogate due to a decrease of projection matrix size. These findings open up the question of utility of the proposed eb-POD method. Is an increase in boundary accuracy worth the decrease in overall groundwater head accuracy? This is a question that can not be answered generally. Instead, the correct answer to this question depends fully on the modeler, the model and the tackled problem. Are fluxes over boundaries important predictions of the model? Where, in time and space, does the increase in general groundwater head error occur through the inclusion of eb-POD? These are some of the questions that a modeler applying POD, and pondering to apply eb-POD, should ask themselves.

Fortunately, eb-POD is very simple to apply, once the setup for general application of POD to the complex model is existing. One just needs to identify the relevant boundary cells and add a few lines of simple linear algebra to the code. Furthermore, since the basic POD setup is necessary for the application of eb-POD, it is very fast and easy to compute surrogate model results with and without the application of eb-POD. This allows the quantification of both the gain in boundary accuracy and the loss in general groundwater head accuracy. On basis of the exact numbers, the modeler can thus make an informed decision about which, if any, boundaries to treat by eb-POD in their model. These advantages make the newly proposed eb-POD a simple yet informative addition to applying POD in groundwater modeling, giving the modeler a quantifiable choice for more precise boundary representation at a minor cost of overall accuracy.

As discussed in this thesis, the estimation of model simplification error is essential for the informed application of surrogate models as predictors. I adapted an uncertainty estimation method to quantify model simplification errors and applied it for the comparison of model predictions of a real-world, complex benchmark models and three different surrogate models.

The analysis of simplification errors suggests that ANN surrogates are fast, suitable predictors in groundwater systems for predictions where sufficient training data is available to inform the ANN model. The chosen test case for structural simplification methods, a spatially and parametrically simplified MODFLOW model,

shows ability to reproduce complex model predictions more certainly the closer they where in location and data type to the calibration data set. Furthermore, predictions in model regions necessitating high spatial and parametrical variability or highly dependent on certain, now over-simplified model parameters, prove to be highly uncertain in the surrogate model. A POD surrogate shows similar behavior to the simplified MODFLOW model regarding the certainty of its model predictions correlating to closeness of said predictions to the calibration data set. In addition, the POD surrogate struggles to believably predict system states dependent on non-linear processes in the complex model due to the underlying linearization of the POD surrogate.

These results prove mainly one thing: the correct use of a surrogate model as a predictor instead of a complex one highly depends on the chosen type of surrogate and model prediction. While this is obviously differing in each specific case, I found certain patterns in the simplification error analysis I deem generalizable. Surrogate models, in general, work best predicting system states that are similar in location and/or data type to the data set used to calibrate them. If certain model structures or workings are omitted in the simplification process, like non-linearities or certain boundaries, predictions depending on these model structures can be assumed to be erroneous in the surrogate models. The same is true where parameter detail gets lost in the simplification. This means model simplification has to be undertaken with a clear vision of the purpose of the surrogate and the predictions made with it. A simplification error analysis like the one presented in this thesis can aid such an aimed simplification process by testing the surrogate's ability to trustworthily reproduce key model predictions.

Finally, I tested two of the three surrogates in their ability to be applied for analysis of worth of existing, "future" and "parametric" data using a fast first-order second-moment method while taking into account parametric uncertainty through null-space Monte-Carlo methods. Due to their lack of physical distribution of outputs, the ANNs were excluded from this analysis.

The structural simplification surrogate could reproduce results regarding the worth of existing data, but not "future" or "parametric" data. Its inability to

estimate worth of "future" data is likely caused by its simplification regarding certain processes and model predictions, so it does show to be somewhat reliable from some predictions. "Parametric" data worth can not be estimated by this surrogate due to its high simplification in parameter space. This could be different for such a structural simplification surrogate which keeps the original parameter space, though the decrease in model run time in application would suffer. The POD surrogate agrees with the complex model regarding worth of all three types of data, irrespective of model prediction. This shows that while it may not be a good predictor in certain cases, as discussed above, it could still be applied to estimate worth of data for such model predictions. The high variance of its data worth results suggests that parametric uncertainty has to be taken into account for such reliable results, though, indicating the necessity of the presented robust data worth analysis.

In general, data worth analysis using surrogate models is possible, as the results show. Correct estimates of data worth depend on the chosen surrogate, the prediction of interest and the type of data that is analyzed, though. Again, surrogates produce more trustworthy estimates of data worth for predictions made with higher certainty by the surrogates, and are best at computing worth of existing measurement data. They seem to be applicable for the estimation of worth of "future" measurements, as well, especially methods like POD that inclusively keep the spatial dimensions of the complex model. The same is true for POD surrogates (and maybe other surrogates keeping the full parameter space) in regard to "parametric" data worth. Results generated via first-order second-moment methods and surrogates have to be taken with caution, though, as the high variance of many data worth estimates by the surrogates suggests that parametric uncertainty has to be taken into account. For this, the proposed method to combine first-order second-moment uncertainty estimation methods with multiple parameter fields efficiently generated by, for example, null-space Monte-Carlo, seems to be sufficient.

6 Conclusions and outlook

From the summary of the research topic of surrogate modeling in groundwater presented in chapter 2, I formulated three distinct objectives for this thesis in chapter 3. I presented methods for tackling these objectives in chapter 4 and discussed the results in chapter 5. Now, I draw some conclusions about each of the three objectives and the overarching topic of surrogates in groundwater and give an outlook about remaining challenges in the field.

The first research question formulated in this thesis was: *"can the implicit error in treating boundary conditions in POD be improved?"*. I presented the eb-POD extension for eliminating (or reducing) errors at boundaries in POD and successfully proved its performance, albeit at a small cost in general groundwater head accuracy of the surrogate. From these findings, I draw the following conclusions:

- eb-POD provides an easy, fast framework for POD modelers to quantify the trade-off between eliminating boundary errors and general accuracy of the surrogate.

- problem-specific surrogate modeling should be based on informed decisions about strengths and weaknesses of the applied surrogate. eb-POD is a tool to help make such informed decisions and thus should be integrated into POD in groundwater as an overall improvement.

The second research question addressed in this thesis was: *"which types of surrogate models are suitable to reproduce different complex model predictions with a small model simplification error?"*. I built three different surrogates from all three major categories on basis of a complex benchmark model and estimated simplification errors of these surrogates for various predictions. From these results, I conclude:

- general performance of a surrogate as predictor depends on the type of surrogate and prediction in question.

- ANNs are highly accurate predictors where sufficient training data pertaining to the prediction is available. Predictions made by both structural sim-

plification methods and projection-based methods get worse with increasing difference to the calibration data set. Structural simplification surrogate predictions are more certain the less dependent on omitted features and spatial or parametrical variance they are. POD surrogates suffer from linearization effects on predictions depending on model non-linearity.

- correct choice of surrogate model type and the degree of simplification in space, parameters and features must thus be made dependent on the prediction of interest. Successful surrogate modeling application needs purposeful, target-oriented construction of surrogates.

The third research question presented was: *"can surrogate models be used to reproduce worth of data estimated by a complex model via a robust uncertainty analysis method?"*. After eliminating the ANNs due to their inability to estimate data worth, I applied the other two surrogates to a robust data worth analysis regarding existing, "future" and "parametric" data while taking parametric uncertainty into account. From the comparison of the results with the complex benchmark model, I draw these conclusions:

- Surrogates of the structural simplification category struggle to estimate worth of "future" and "parametric" data due to lack of necessary spatial and parametrical resolution. Projection-based surrogates retain the original spatial and parametric dimensions and prove successful in estimating worth of existing, "future" and "parametric" data.

- the high variance associated with many surrogate data worth estimations suggests the necessity of an extension of first-order, second-moment uncertainty analysis methods to account for parametric uncertainty.

Overall, this work has effectively compared three different surrogate models regarding their strengths and weaknesses for predictive purposes and data worth estimation. The main conclusion drawn for all of the three research questions posed in this thesis is that surrogate models can be applied successfully as predictors and estimators of data worth in groundwater. Nonetheless, simplification in

spatial and parametric resolution, as well as general features, highly affects certain surrogate predictions dependent on those features. Thus, responsible application of surrogates requires a purposeful balancing of the surrogate and its setup for the target question.

The methods presented in this thesis can help researchers to identify strengths and weaknesses of surrogate models regarding predictions of interest. This could be used to iteratively improve surrogate models for their specific applications. This is very time-consuming, though, thus undermining the advantage of surrogates in general. Therefore, more research to identify best practice and guidelines for target-oriented surrogate modeling needs to be done. While POD surrogates arguably performed best in regards to overall use in this thesis, they are still hampered by model non-linearity. Certain advancements have been made in this regard, but it is by no means a finished topic. Furthermore, POD application is highly complex, demonstrating the need for widely available implementation into existing groundwater modeling software. The advances in application of surrogate models for data worth analysis made in this thesis suggest their potential use in network design and optimization of cost/use analysis as well.

References

Adamowski, J., Chan, H. F., 2011. A wavelet neural network conjunction model for groundwater level forecasting. Journal of Hydrology 407 (1), 28 – 40. https://doi.org/10.1016/j.jhydrol.2011.06.013

Asher, M. J., Croke, B. F. W., Jakeman, A. J., Peeters, L. J. M., 2015. A review of surrogate models and their application to groundwater modeling. Water Resources Research 51 (8), 5957–5973. https://doi.org/10.1002/2015WR016967

Beven, K., Binley, A., 1992. The future of distributed models: Model calibration and uncertainty prediction. Hydrological Processes 6 (3), 279–298. https://doi.org/10.1002/hyp.3360060305

Boyce, S. E., Nishikawa, T., Yeh, W. W.-G., 2015. Reduced order modeling of the newton formulation of modflow to solve unconfined groundwater flow. Advances in Water Resources 83, 250 – 262. https://doi.org/10.1016/j.advwatres.2015.06.005

Boyce, S. E., Yeh, W. W.-G., 2014. Parameter-independent model reduction of transient groundwater flow models: Application to inverse problems. Advances in Water Resources 69, 168–180. https://doi.org/10.1016/j.advwatres.2014.04.009

Burrows, W., Doherty, J., 2015. Efficient calibration/uncertainty analysis using paired complex/surrogate models. Groundwater 53 (4), 531–541. https://doi.org/10.1111/gwat.12257

Burrows, W., Doherty, J., 2016. Gradient-based model calibration with proxy-model assistance. Journal of Hydrology 533, 114 – 127. https://doi.org/10.1016/j.jhydrol.2015.11.033

Cardoso, M. A., Durlofsky, L. J., Sarma, P., 2009. Development and application of reduced-order modeling procedures for subsurface flow simulation. In-

ternational Journal for Numerical Methods in Engineering 77 (9), 1322–1350. https://doi.org/10.1002/nme.2453

Cirpka, O. A., Bürger, C. M., Nowak, W., Finkel, M., 2004. Uncertainty and data worth analysis for the hydraulic design of funnel-and-gate systems in heterogeneous aquifers. Water Resources Research 40, W11502. https://doi.org/10.1029/2004WR003352

Cooley, R. L., Christensen, S., 2006. Bias and uncertainty in regression-calibrated models of groundwater flow in heterogeneous media. Advances in Water Resources 29 (5), 639 – 656. https://doi.org/10.1016/j.advwatres.2005.07.012

Daliakopoulos, I., Coulibaly, P., K. Tsanis, I., 2005. Groundwater level forecasting using artificial neural networks. Journal of Hydrology 309, 229–240. https://doi.org/10.1016/j.jhydrol.2004.12.001

Darcy, H., 1856. Les fontaines publiques de la ville de Dijon. Exposition et application des principes à suivre et des formules à employer dans les questions de distribution d'eau: ouvrage terminé par un appendice relatif aux fournitures d'eau de plusieurs villes au filtrage des eaux et à la fabrication des tuyaux de fonte, de plomb, de tole et de bitume. Dalmont.

Dausman, A. M., Doherty, J., Langevin, C. D., Sukop, M. C., 2010. Quantifying data worth toward reducing predictive uncertainty. Groundwater 48 (5), 729–740. https://doi.org/10.1111/j.1745-6584.2010.00679.x

Doherty, J., 2016. PEST: Model-Independent Parameter Estimation - User Manual. Watermark Numerical Computing, 6th Edition.

Doherty, J., Christensen, S., 2011. Use of paired simple and complex models to reduce predictive bias and quantify uncertainty. Water Resources Research 47 (12), W12534. https://doi.org/10.1029/2011WR010763

Erickson, M., Mayer, A., Horn, J., 2002. Multi-objective optimal design of groundwater remediation systems: application of the niched pareto genetic algorithm (npga). Advances in Water Resources 25 (1), 51 – 65. https://doi.org/10.1016/S0309-1708(01)00020-3

Feyen, L., Gorelick, S. M., 2005. Framework to evaluate the worth of hydraulic conductivity data for optimal groundwater resources management in ecologically sensitive areas. Water Resources Research 41 (3). https://doi.org/10.1029/2003WR002901

Fienen, M., Doherty, J., Hunt, R., Reeves, H., 2010. Using prediction uncertainty analysis to design hydrologic monitoring networks: Example applications from the Great Lakes water availability pilot project. Tech. Rep. 5159, U.S. Geological Survey Scientific Investigations Report.

Fienen, M. N., Masterson, J. P., Plant, N. G., Gutierrez, B. T., Thieler, E. R., 2013. Bridging groundwater models and decision support with a Bayesian network. Water Resources Research 49 (10), 6459–6473. https://doi.org/10.1002/wrcr.20496

Fu, J., Gómez-Hernández, J. J., 2009. Uncertainty assessment and data worth in groundwater flow and mass transport modeling using a blocking Markov chain Monte Carlo method. Journal of Hydrology 364 (3), 328 – 341. https://doi.org/10.1016/j.jhydrol.2008.11.014

Galerkin, B. G., 1968. Rods and plates: series in some questions of elastic equilibrium of rods and plates. National Technical Information Service.

Gleick, P. H. (Ed.), 1993. Water in Crisis: A Guide to the World's Fresh Water Resources. Oxford University Press, New York.

Gosses, M., Nowak, W., Wöhling, T., 2018. Explicit treatment for Dirichlet, Neumann and Cauchy boundary conditions in POD-based reduction of groundwater models. Advances in Water Resources 115, 160 – 171. https://doi.org/10.1016/j.advwatres.2018.03.011

Gosses, M., Wöhling, T., 2019. Simplification error analysis for groundwater predictions with reduced order models. Advances in Water Resources 125, 41–56. https://doi.org/10.1016/j.advwatres.2019.01.006

Gunzburger, M. D., Peterson, J. S., Shadid, J. N., 2007. Reduced-order modeling of time-dependent PDEs with multiple parameters in the boundary data. Computer Methods in Applied Mechanics and Engineering 196, 1030 – 1047. https://doi.org/10.1016/j.cma.2006.08.004

Hall, J., Solomatine, D., 2008. A framework for uncertainty analysis in flood risk management decisions. International Journal of River Basin Management 6 (2), 85–98. https://doi.org/10.1080/15715124.2008.9635339

Hill, M. C., Tiedeman, C. R., 2006. Effective groundwater model calibration: with analysis of data, sensitivities, predictions, and uncertainty. John Wiley & Sons.

James, A. L., Oldenburg, C. M., 1997. Linear and Monte Carlo uncertainty analysis for subsurface contaminant transport simulation. Water Resources Research 33 (11), 2495–2508. https://doi.org/10.1029/97WR01925

Kavetski, D., Kuczera, G., Franks, S. W., 2006. Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. Water Resources Research 42 (3), W03407. https://doi.org/10.1029/2005wr004368

Keating, E. H., Doherty, J., Vrugt, J. A., Kang, Q., 2010. Optimization and uncertainty assessment of strongly nonlinear groundwater models with high parameter dimensionality. Water Resources Research 46 (10), W10517. https://doi.org/10.1029/2009WR008584

Khu, S.-T., Werner, M. G. F., 2003. Reduction of Monte-Carlo simulation runs for uncertainty estimation in hydrological modelling. Hydrology and Earth System Sciences Discussions 7 (5), 680–692.

Kunstmann, H., Kinzelbach, W., Siegfried, T., 2002. Conditional first-order second-moment method and its application to quantification of uncer-

tainty in groundwater modeling. Water Resources Research 38 (4), W1035. https://doi.org/10.1029/2000WR000022

Laloy, E., Rogiers, B., Vrugt, J. A., Mallants, D., Jacques, D., 2013. Efficient posterior exploration of a high-dimensional groundwater model from two-stage Markov chain Monte Carlo simulation and polynomial chaos expansion. Water Resources Research 49 (5), 2664–2682. https://doi.org/10.1002/wrcr.20226

Marzouk, Y. M., Najm, H. N., 2009. Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems. Journal of Computational Physics 228 (6), 1862 – 1902. https://doi.org/10.1016/j.jcp.2008.11.024

Matott, L. S., Babendreier, J. E., Purucker, S. T., 2009. Evaluating uncertainty in integrated environmental models: a review of concepts and tools. Water Resources Research 45 (6). https://doi.org/10.1029/2008WR007301

McPhee, J., Yeh, W. W.-G., 2008. Groundwater management using model reduction via empirical orthogonal functions. Journal of Water Resources Planning and Management 134 (2), 161–170. https://doi.org/10.1061/(ASCE)0733-9496(2008)134:2(161)

Moore, C., Doherty, J., 2006. The cost of uniqueness in groundwater model calibration. Advances in Water Resources 29 (4), 605 – 623. https://doi.org/10.1016/j.advwatres.2005.07.003

Mulligan, K. B., Ahlfeld, D. P., 2016. Model reduction for combined surface water/groundwater management formulations. Environmental Modelling & Software 81, 102 – 110. https://doi.org/10.1016/j.envsoft.2016.03.013

Najm, H. N., 2009. Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics. Annual Review of Fluid Mechanics 41 (1), 35–52. https://doi.org/10.1146/annurev.fluid.010908.165248

Niswonger, R., Panday, S., Ibaraki, M., 2011. MODFLOW-NWT, a Newton formulation for MODFLOW-2005. Tech. Rep. 6-A37, U.S. Geological Survey Techniques and Methods.

Niswonger, R., Prudic, D., 2005. Documentation of the Stream flow-Routing (SFR2) package to include unsaturated flow beneath streams - a modification to SFR1. Tech. Rep. 6-A13, U.S. Geological Survey Techniques and Methods.

Pappenberger, F., Beven, K. J., 2006. Ignorance is bliss: Or seven reasons not to use uncertainty analysis. Water Resources Research 42 (5). https://doi.org/10.1029/2005WR004820

Pasetto, D., Guadagnini, A., Putti, M., 2011. POD-based Monte Carlo approach for the solution of regional scale groundwater flow driven by randomly distributed recharge. Advances in Water Resources 34 (11), 1450–1463. https://doi.org/10.1016/j.advwatres.2011.07.003

Pasetto, D., Putti, M., Yeh, W. W.-G., 2013. A reduced-order model for groundwater flow equation with random hydraulic conductivity: Application to Monte Carlo methods. Water Resources Research 49 (6), 3215–3228. https://doi.org/10.1002/wrcr.20136

Razavi, S., Tolson, B. A., Burn, D. H., 2012a. Numerical assessment of metamodelling strategies in computationally intensive optimization. Environmental Modelling & Software 34, 67 – 86. https://doi.org/10.1016/j.envsoft.2011.09.010

Razavi, S., Tolson, B. A., Burn, D. H., 2012b. Review of surrogate modeling in water resources. Water Resources Research 48 (7), W07401. https://doi.org/10.1029/2011wr011527

Refsgaard, J. C., Christensen, S., Sonnenborg, T. O., Seifert, D., Højberg, A. L., Troldborg, L., 2012. Review of strategies for handling geological uncertainty in groundwater flow and transport modeling. Advances in Water Resources

36, 36 – 50, special Issue on Uncertainty Quantification and Risk Assessment. https://doi.org/10.1016/j.advwatres.2011.04.006

Refsgaard, J. C., Storm, B., 1996. Construction, Calibration And Validation of Hydrological Models. Springer Netherlands, Dordrecht, pp. 41–54. https://doi.org/10.1007/978-94-009-0257-23

Schöniger, A., Illman, W. A., Wöhling, T., Nowak, W., 2015. Finding the right balance between groundwater model complexity and experimental effort via Bayesian model selection. Journal of Hydrology 531, 96 – 110. https://doi.org/10.1016/j.jhydrol.2015.07.047

Schultz, M. T., Small, M. J., Fischbeck, P. S., Farrow, R. S., 2006. Evaluating response surface designs for uncertainty analysis and prescriptive applications of a large-scale water quality model. Environ. Model. Assess. 11(4), 345–359. https://doi.org/10.1007/s10666-006-9043-9

Siade, A. J., Hall, J., Karelse, R. N., 2017. A practical, robust methodology for acquiring new observation data using computationally expensive groundwater models. Water Resources Research 53 (11), 9860–9882. https://doi.org/10.1002/2017WR020814

Siade, A. J., Putti, M., Yeh, W. W.-G., 2010. Snapshot selection for groundwater model reduction using proper orthogonal decomposition. Water Resources Research 46 (8), W08539. https://doi.org/10.1029/2009WR008792

Siade, A. J., Putti, M., Yeh, W. W.-G., 2012. Reduced order parameter estimation using quasilinearization and quadratic programming. Water Resources Research 48 (6), W06502. https://doi.org/10.1029/2011WR011471

Stanko, Z. P., Boyce, S. E., Yeh, W. W.-G., 2016. Nonlinear model reduction of unconfined groundwater flow using pod and deim. Advances in Water Resources 97, 130 – 143. https://doi.org/10.1016/j.advwatres.2016.09.005

Strang, G., 2016. Introduction to Linear Algebra, Fifth Edition. Wellesley-Cambridge Press.

Sun, A. Y., 2008. State-of-the-art multiscale approaches for flow and transport modeling: a literature review. Tech. rep., Center for Nuclear Waste Regulatory Analyses, San Antonio, Texas.

Tait, A., R., H., R., T., X.G., Z., 2006. Thin plate smoothing spline interpolation of daily rainfall for New Zealand using a climatological rainfall surface. International Journal of Climatology 26 (14), 2097 – 2115. https://doi.org/10.1002/joc.1350

Taormina, R., wing Chau, K., Sethi, R., 2012. Artificial neural network simulation of hourly groundwater levels in a coastal aquifer system of the Venice lagoon. Engineering Applications of Artificial Intelligence 25 (8), 1670 – 1676. https://doi.org/10.1016/j.engappai.2012.02.009

Ushijima, T. T., Yeh, W. W.-G., 2015. Experimental design for estimating unknown hydraulic conductivity in an aquifer using a genetic algorithm and reduced order model. Advances in Water Resources 86, 193 – 208. https://doi.org/10.1016/j.advwatres.2015.09.029

Vermeulen, P., Heemink, A., Stroet, C. T., 2004a. Reduced models for linear groundwater flow models using empirical orthogonal functions. Advances in Water Resources 27 (1), 57–69. https://doi.org/10.1016/j.advwatres.2003.09.008

Vermeulen, P. T. M., Heemink, A. W., te Stroet, C. B. M., 2004b. Low-dimensional modelling of numerical groundwater flow. Hydrological Processes 18 (8), 1487–1504. https://doi.org/10.1002/hyp.1424

Vermeulen, P. T. M., te Stroet, C. B. M., Heemink, A. W., 2006. Model inversion of transient nonlinear groundwater flow models using model reduction. Water Resources Research 42 (9), W09417. https://doi.org/10.1029/2005WR004536

von Gunten, D., Wöhling, T., Haslauer, C., Merchán, D., Causapé, J., Cirpka, O., 2014. Efficient calibration of a distributed pde-based hydrolog-

ical model using grid coarsening. Journal of Hydrology 519, 3290 – 3304. https://doi.org/10.1016/j.jhydrol.2014.10.025

Watson, T. A., Doherty, J. E., Christensen, S., 2013. Parameter and predictive outcomes of model simplification. Water Resources Research 49 (7), 3952–3977. https://doi.org/10.1002/wrcr.20145

Willcox, K., Megretski, A., 2005. Fourier series for accurate, stable, reduced-order models in large-scale linear applications. SIAM Journal on Scientific Computing 26 (3), 944–962. https://doi.org/10.1137/S1064827502418768

Winston, R., 2009. ModelMuse - a graphical user interface for MODFLOW-2005 and PHAST. Tech. Rep. 6-A29, U.S. Geological Survey Techniques and Methods.

Winton, C., Pettway, J., Kelley, C., Howington, S., Eslinger, O. J., 2011. Application of Proper Orthogonal Decomposition (POD) to inverse problems in saturated groundwater flow. Advances in Water Resources 34 (12), 1519 – 1526. https://doi.org/10.1016/j.advwatres.2011.09.007

Wöhling, T., Geiges, A., Nowak, W., 2016. Optimal Design of Multitype Groundwater Monitoring Networks Using Easily Accessible Tools. Groundwater 54 (6), 861–870. https://doi.org/10.1111/gwat.12430

Wöhling, T., Gosses, M. J., Wilson, S. R., Davidson, P., 2018. Quantifying river-groundwater interactions of New Zealand's gravel-bed rivers: The Wairau Plain. Groundwater 56, 647 – 666. https://doi.org/10.1111/gwat.12625

Woodbury, A. D., Dunbar, W. S., Nour-Omid, B., 1990. Application of the Arnoldi Algorithm to the solution of the advection-dispersion equation. Water Resources Research 26 (10), 2579–2590. https://doi.org/10.1029/wr026i010p02579

WWAP (UNESCO World Water Assessment Programme), 2019. The United Nations World Water Development Report 2019: Leaving No One Behind. Paris, UNESCO.

Yoon, H., Jun, S.-C., Hyun, Y., Bae, G.-O., Lee, K.-K., 2011. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. Journal of Hydrology 396 (1), 128 – 138. https://doi.org/10.1016/j.jhydrol.2010.11.002

A Appendix: Publications

1. Gosses, M., Nowak, W., Wöhling, T., 2018. Explicit treatment for Dirichlet, Neumann and Cauchy boundary conditions in POD-based reduction of groundwater models. Advances in Water Resources 115, 160–171. https://doi.org/10.1016/j.advwatres.2018.03.011

2. Gosses, M., Wöhling, T., 2019. Simplification error analysis for groundwater predictions with reduced order models. Advances in Water Resources 125, 4156. https://doi.org/10.1016/j.advwatres.2019.01.006

Advances in Water Resources 115 (2018) 160–171

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

Explicit treatment for Dirichlet, Neumann and Cauchy boundary conditions in POD-based reduction of groundwater models

Moritz Gosses[a,*], Wolfgang Nowak[b], Thomas Wöhling[a,c]

[a] Department of Hydrology, Technische Universität Dresden, Dresden 01069, Germany
[b] Institute of Hydraulic Engineering, SimTech, University of Stuttgart, Stuttgart 70569, Germany
[c] Lincoln Agritech Ltd., Ruakura Research Centre, Hamilton 3240, New Zealand

ARTICLE INFO

Article history:
Received 9 August 2017
Revised 14 March 2018
Accepted 15 March 2018
Available online 16 March 2018

Keywords:
Model reduction
Proper orthogonal decomposition
Boundary condition treatment

ABSTRACT

In recent years, proper orthogonal decomposition (POD) has become a popular model reduction method in the field of groundwater modeling. It is used to mitigate the problem of long run times that are often associated with physically-based modeling of natural systems, especially for parameter estimation and uncertainty analysis. POD-based techniques reproduce groundwater head fields sufficiently accurate for a variety of applications. However, no study has investigated how POD techniques affect the accuracy of different boundary conditions found in groundwater models. We show that the current treatment of boundary conditions in POD causes inaccuracies for these boundaries in the reduced models. We provide an improved method that splits the POD projection space into a subspace orthogonal to the boundary conditions and a separate subspace that enforces the boundary conditions. To test the method for Dirichlet, Neumann and Cauchy boundary conditions, four simple transient 1D-groundwater models, as well as a more complex 3D model, are set up and reduced both by standard POD and POD with the new extension. We show that, in contrast to standard POD, the new method satisfies both Dirichlet and Neumann boundary conditions. It can also be applied to Cauchy boundaries, where the flux error of standard POD is reduced by its head-independent contribution. The extension essentially shifts the focus of the projection towards the boundary conditions. Therefore, we see a slight trade-off between errors at model boundaries and overall accuracy of the reduced model. The proposed POD extension is recommended where exact treatment of boundary conditions is required.

1. Introduction

The building and application of groundwater models is one of the major fields supporting modern groundwater management. As these models often couple well-understood physical principles with insufficiently known and highly spatially variable aquifer properties, computerized calibration of parameters and robust uncertainty analysis of model predictions are mandatory for meaningful interpretation of the results. Unfortunately, calibration and uncertainty analysis often require many model runs. Modern advances in computer performance and parallelization can alleviate these run times only to a limited extent.

Over the years, many researchers tackled this problem with model reduction. Many different methods have been applied to reduce run times of physically-based groundwater models. Usually, model (run-time) reduction comes at a cost of model accuracy. Balancing this trade-off has been the main focus of research. Model reduction techniques in groundwater can be roughly put into three categories: data-driven methods, projection-based methods and structural reduction methods. The data-driven methods include artificial neural networks (Taormina et al., 2012), certain forms of polynomial chaos expansion (Oladyshkin and Nowak, 2012) or Bayesian networks (Fienen et al., 2013). The most prominent projection-based method is proper orthogonal decomposition (POD) (Vermeulen et al., 2004b; McPhee and Yeh, 2008 or Siade et al., 2010, for example), followed by the Fourier model reduction (Willcox and Megretski, 2005). Structural reduction methods directly reduce the size of the model discretization or simplify the representation of the processes (von Gunten et al., 2014). Structural simplification can go along with parameter reduction, like Doherty and Christensen (2011) did with the inversion-based upscaling approach.

Proper orthogonal decomposition projects the discretized model equations onto a subspace, usually by the Galerkin method. The subspace is mostly defined via singular vectors of simulations of

* Corresponding author.
E-mail address: moritz.gosses@tu-dresden.de (M. Gosses).

https://doi.org/10.1016/j.advwatres.2018.03.011
0309-1708/

the full model for pre-defined times and forcings, so-called snapshots (see Winton et al., 2011 for an example of POD in groundwater not based on snapshots). Recent research in POD for groundwater models focused on non-linear reduction (e.g. Stanko et al., 2016), snapshot selection (e.g. Siade et al., 2010) or inverse modeling (Boyce and Yeh, 2014 or Ushijima and Yeh, 2015, for example), but did not focus on the effects of POD on the boundary conditions of the system. The full physically-based groundwater model depends on the statement of and adherence to certain boundary conditions. While the POD method incorporates the original model equations, it does not explicitly adhere to the original boundary conditions. The approximation of groundwater heads in the reduced-order model (ROM) obtained by POD may lead to significant errors in boundary heads or fluxes. This is especially relevant if the ROM is used to predict boundary fluxes (groundwater discharge into rivers or low-land springs, for example).

This problem has been analyzed in other research areas. Gunzburger et al. (2007) developed a method to treat Dirichlet boundaries explicitly by modifying the POD subspace to be orthogonal to the boundary conditions. This leads to an underdetermined system of equations that is then filled up with the corresponding projections of the original boundary equations. Cosimo et al. (2016) built on this work by proposing different methods of building the ROM, but still only with respect to Dirichlet boundaries. We transfer the idea proposed by Gunzburger et al. (2007) into the area of groundwater modeling. Furthermore, we significantly extend the method, now allowing the explicit treatment of all three types of boundary conditions (Dirichlet, Neumann and Cauchy) in POD-reduction of groundwater models. We analyze the performance of the new method with several test cases (Section 3) and study the trade-off between overall model performance and accuracy at the boundary conditions (Section 4).

2. Methods

In this section, we will first give a short summary on the mathematical basis of the groundwater models to which POD is typically applied. Second, we briefly revisit the common POD method. We will then develop the extension of POD for explicit-boundary treatment (eb-POD) and explain the implications of its use. Finally, we will present different synthetic models used in this study to depict and test the new method, as well as the chosen methodology for snapshot selection.

2.1. (Full) groundwater model

In this work, we focus on groundwater models based on the groundwater flow equations. These equations have been presented in numerous studies (Pinder and Celia, 2006). Thus we only summarized them here as required for developing the new method. More in-depth formulations can be found in the MODFLOW manual (Harbaugh, 2005), for example.

3D groundwater flow is usually described by the following partial differential equation, based on Darcy's (1856) law:

$$\frac{\partial}{\partial x}\left(K_{xx}\frac{\partial h}{\partial x}\right)+\frac{\partial}{\partial y}\left(K_{yy}\frac{\partial h}{\partial y}\right)+\frac{\partial}{\partial z}\left(K_{zz}\frac{\partial h}{\partial z}\right)+W=S\frac{\partial h}{\partial t} \tag{1}$$

with K being hydraulic conductivity [L/T], h the piezometric head [L], W a source/sink term [1/T] and S the storage term [1/L].

The solution of this differential equation requires initial and boundary conditions. Initial conditions are defined by specifying the head $h(x,y,z,0)=f_0(x,y,z)$ at time zero. For the boundary conditions, three different types are most common:

$$\begin{array}{ll} h=f_1(t,x,y,z) & \text{Dirichlet (Type I)} \\ \vec{n}\cdot\nabla h=f_2(t,x,y,z) & \text{Neumann (Type II)} \\ \vec{n}\cdot\nabla h+c(t,x,y,z)h=f_3(t,x,y,z) & \text{Cauchy (Type III)} \end{array} \tag{2}$$

where $\nabla=(\frac{\partial}{\partial x},\frac{\partial}{\partial y},\frac{\partial}{\partial z})$, $\vec{n}$ is the normal vector to the boundary and $f_{1,2,3}$ and c can be functions in space and/or time. Conceptually, Dirichlet boundary conditions represent specified heads, Neumann boundaries are specified fluxes, and Cauchy boundaries are specified leakages.

Spatial and temporal discretization of Eq. (1) and its boundary conditions (e.g. by Finite Differences, Finite Volumes or Finite Elements) leads to the following linear system of equations, provided here for implicit time-stepping in matrix notation:

$$(\mathbf{R}-\mathbf{L}\triangle t)\mathbf{h}^{t+1}=\mathbf{R}\mathbf{h}^{t}+\triangle t\mathbf{q} \tag{3}$$

where $\mathbf{R}=f(S)$, $\mathbf{L}=f(K)$ and $\mathbf{q}=f(W)$.

After renaming for convenience, $(\mathbf{R}-\mathbf{L}\triangle t)=\mathbf{A}$ and $\mathbf{R}\mathbf{h}^{t}+\triangle t\mathbf{q}=\mathbf{b}$, Eq. (3) becomes:

$$\mathbf{A}\mathbf{h}^{t+1}=\mathbf{b} \tag{4}$$

This system of linear equations is solved for every time step. The size of the system of equations depends on the number of model cells, n_m, as follows: $\mathbf{A}\in\mathbb{R}^{n_m\times n_m}$ and $\mathbf{b},\mathbf{h}\in\mathbb{R}^{n_m\times 1}$. The boundary conditions from Eq. (2) can be found explicitly in boundary-related rows of the system in Eq. (4). The solution of the equations for groundwater levels $\mathbf{h}$ requires iterative manipulation of $\mathbf{A}$, a computationally (and therefore time-) demanding process.

2.2. Proper orthogonal decomposition (POD)

The POD method aims to reduce the size of Eq. (4) by projection onto a suitable subspace. This subspace is created by singular value decomposition of a collection of previously collected system states (so-called snapshots).

These snapshots are built by calculating the vectors of groundwater heads $\mathbf{h}$ at specific times $\mathbf{t}_s=[t_{s1},t_{s2},\ldots,t_{sn_s}]$ in the full model and saving them in the matrix $\mathbf{H_s}=[\mathbf{h}_{s1},\mathbf{h}_{s2},\ldots,\mathbf{h}_{sn_s}]$. The number of snapshots n_s results in $\mathbf{H_s}\in\mathbb{R}^{n_m\times n_s}$. The process of snapshot selection is detailed further in Section 2.5.

Now, the snapshot matrix $\mathbf{H}_s$ is treated by singular value decomposition,

$$\mathbf{H_s}=\mathbf{U}\mathbf{\Sigma}\mathbf{V}^{T} \tag{5}$$

resulting in the matrix of left singular vectors $\mathbf{U}\in\mathbb{R}^{n_m\times n_m}$, the diagonal matrix of singular values $\mathbf{\Sigma}\in\mathbb{R}^{n_m\times n_s}$ and the matrix of right singular vectors $\mathbf{V}\in\mathbb{R}^{n_s\times n_s}$. The projection matrix $\mathbf{P}=[\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_{n_r}]$ is a selection of left singular vectors from $\mathbf{U}$, using the singular values σ as selection criteria, similar to a principal component analysis. The number of retained singular values, n_r, is chosen as follows:

$$\frac{\sum_{i=1}^{n_r}\sigma_i}{\sum_{i=1}^{rank(S)}\sigma_i}\times 100\geq \mathrm{RV} \tag{6}$$

where RV is a chosen specific relative retained variance (i.e. of 99.99%). Since singular values are ranked in descending order, this assures that the chosen RV of the snapshot set is achieved with the lowest possible number of singular vectors. n_r is the size of the reduced model.

Now, the Galerkin (1968) method is used to project Eq. (4) onto the subspace spanned by the columns of $\mathbf{P}$:

$$\mathbf{P}^{T}\mathbf{A}\mathbf{P}\mathbf{r}^{t+1}=\mathbf{P}^{T}\mathbf{b} \tag{7}$$

Solving Eq. (7) is much more time-efficient compared to the original Eq. (4) because $n_m\gg n_r$. The groundwater heads in the

original model are calculated from this as follows:

$$\mathbf{h} \approxeq \mathbf{Pr} \tag{8}$$

This approximation captures the RV% of the dynamics represented by the snapshots in $\mathbf{H_s}$. However, no explicit lines can be found anymore that enforce the boundary conditions.

2.3. Explicit-boundary POD

2.3.1. General form

The original Eq. (4) handles boundary conditions explicitly to assure the correct solution in affected cells. This does not hold for the reduced model in Eq. (7), where boundary conditions are only approximated due to the projection onto the subspace. We propose an explicit-boundary POD method (eb-POD), which modifies the original POD equations to enforce the boundary conditions in the reduced model. The simplified handling of Dirichlet boundaries (Section 2.3.2) is taken directly from Gunzburger et al. (2007). In this contribution, we generalize their approach for the application to all three types of boundaries (see Eq. (2)), which is subsequently presented.

In our new method, we replace part of the reduced model with the equations for the original boundary conditions. For this, we need to modify the projection matrix $\mathbf{P}$ of the reduced model to remove the approximation regarding these boundaries.

First, we collect the rows of $\mathbf{P}$ corresponding to the independent boundary conditions (bc) we want to treat explicitly into $\mathbf{N}$. Independent boundaries means that for several rows (and therefore model cells) that are all subjected to the same forcing (for example by representing the same constant head boundary along a model edge), only one row has to be collected into $\mathbf{N}$:

$$\mathbf{N} = \mathbf{P}(bc, \cdot) \tag{9}$$

In our notation, the first index of a matrix stands for the position of rows, while the second index stands for the position of columns. $\mathbf{P}(bc, \cdot)$ therefore denotes a subset of $\mathbf{P}$ with the bc rows and all columns, respectively. Likewise, $\mathbf{Q}(\cdot, n_{bc}+1{:}n_r)$ denotes a subset of $\mathbf{Q}$ of all rows and $(n_{bc}+1)$ to n_r columns.

The transpose of matrix $\mathbf{N}$ is now decomposed by QR-factorization:

$$\mathbf{Q}, \mathbf{R} = qr(\mathbf{N}^T) \tag{10}$$

The columns of the resulting matrix $\mathbf{Q}$ of size $n_r \times n_r$ form an orthonormal basis for $\mathbf{N}$. The number of independent boundary conditions n_{bc} must be smaller than the reduced model size n_r. Then, the $(n_{bc}+1) : n_r$ columns of $\mathbf{Q}$ hold the coefficients of the linear combinations of $\mathbf{P}$ that are independent of the boundaries, as the $1{:}\ n_{bc}$ columns of $\mathbf{Q}$ retain all information regarding the boundaries.

Multiplication of these $(n_{bc}+1) : n_r$ columns of $\mathbf{Q}$ with $\mathbf{P}$ leads to:

$$\mathbf{\Psi} = \mathbf{PQ}(\cdot, n_{bc}+1{:}n_r) \tag{11}$$

The matrix $\mathbf{\Psi}$ has the size $n_m \times n_{r2}$, where $n_{r2} = n_r - n_{bc}$. It contains linear combinations of the POD basis functions that vanish at the boundaries. Thus, $\mathbf{\Psi}$ is a reduced version of $\mathbf{P}$ that is cleaned from the inaccurate boundary-related information contained in the original projection matrix. We can now substitute Eq. (11) into Eq. (7):

$$\mathbf{\Psi}^T\mathbf{AP}\mathbf{r}^{t+1} = \mathbf{\Psi}^T\mathbf{b} \tag{12}$$

Because the size n_r of $\mathbf{P}$ and $\mathbf{r}$ is larger by n_{bc} than the number of columns n_{r2} of $\mathbf{\Psi}$, Eq. (12) is underdetermined. As $\mathbf{\Psi}$ is boundary-independent and the number of missing equations is n_{bc}, we can now add the accurate boundary conditions from the original model:

$$\mathbf{A}(bc, \cdot)\mathbf{h}^{t+1} = \mathbf{b}(bc) \tag{13}$$

As Eq. (12) works with the reduced solution $\mathbf{r}$ instead of $\mathbf{h}$, we use the same approximation (Eq. (8)) in Eq. (13):

$$\mathbf{A}(bc, \cdot)\mathbf{P}\mathbf{r}^{t+1} = \mathbf{b}(bc) \tag{14}$$

Combining the new reduced model (Eq. (12)) with the reduced form of the boundary (Eq. (14)) leads to the final system of equations for the eb-POD reduced model:

$$\begin{bmatrix} \mathbf{\Psi}^T\mathbf{AP} \\ \mathbf{A}(bc, \cdot)\mathbf{P} \end{bmatrix} \mathbf{r}^{t+1} = \begin{bmatrix} \mathbf{\Psi}^T\mathbf{b} \\ \mathbf{b}(bc) \end{bmatrix} \tag{15}$$

where the approximation of groundwater heads is again $\mathbf{h} \approxeq \mathbf{Pr}$.

2.3.2. Dirichlet boundaries

Due to the way Dirichlet boundaries are implemented into groundwater models, their handling with the eb-POD simplifies the Eq. (15) for Dirichlet boundary cells (dc). The corresponding system matrix diagonals, $\mathbf{A}(dc, dc)$, are 1, while the off-diagonal elements are 0. This leads to $\mathbf{h}^{t+1}(dc) = \mathbf{b}(dc)$ where $\mathbf{b}(dc)$ is the fixed groundwater head at the Dirichlet boundary. This simplifies $\mathbf{A}(bc, \cdot)\mathbf{P}$ of Eq. (15_ for the Dirichlet boundaries, leading to:

$$\begin{bmatrix} \mathbf{\Psi}^T\mathbf{AP} \\ \mathbf{A}(nc, \cdot)\mathbf{P} \\ \mathbf{P}(dc, \cdot) \end{bmatrix} \mathbf{r}^{t+1} = \begin{bmatrix} \mathbf{\Psi}^T\mathbf{b} \\ \mathbf{b}(nc) \\ \mathbf{b}(dc) \end{bmatrix} \tag{16}$$

where nc denotes the explicitly treated boundaries that are not of Dirichlet-type (Neumann and Cauchy boundaries) and dc represents the Dirichlet boundary cells. If only Dirichlet boundaries are to be handled by eb-POD, Eq. (16) simplifies to the method presented in the field of Applied Mechanical Engineering by Gunzburger et al. (2007). Therefore, Eq. (16) is the generalization of that method to the three featured boundary condition types. Again, the approximation of groundwater heads is still $\mathbf{h} \approxeq \mathbf{Pr}$.

Note that Dirichlet boundaries can potentially be handled correctly by standard POD with several restrictions to the ROM building process:

- the model must simulate drawdown instead of groundwater heads. For drawdown, Dirichlet boundary conditions are always zero, thus not affected by the subspace projection of POD, as has been shown by Siade et al. (2010). Groundwater heads and fluxes can be calculated from the drawdown simulations by simple superposition of drawdown with the groundwater heads without pumping, which are subsequently termed “pseudo-steady-state”. However, this can only be applied to linear groundwater models.
- Dirichlet boundary cells must be omitted in the calculation of the POD ROM (by excluding its rows from the snapshot matrix $\mathbf{H}_s$ and thus from $\mathbf{P}$). Instead these cells are set to the correct values manually. This is sometimes done in groundwater POD in general (presumably for better accuracy at the Dirichlet boundaries), as is stated in Vermeulen et al. (2004a).
- preprocessing of the snapshots is necessary (see Section 2.5) to compute drawdowns by subtraction of the pseudo steady-state. For variable Dirichlet boundaries, this pseudo steady-state heads are time-dependent and have to be computed for every time step. Thus, an additional full model run without pumping is necessary offline to create the corresponding pseudo steady-state. The additional computational requirements reduce the efficiency of the POD model reduction.These limitations in the handling of Dirichlet boundaries are overcome by our proposed eb-POD method.

2.3.3. Neumann boundaries

Neumann boundaries are directly handled by eb-POD as shown in Section 2.3.1. Eb-POD allows the explicit treatment of the Neumann boundaries in the ROM. This ensures the correct boundary

fluxes in the ROM, in contrast to standard POD, where they are approximated.

2.3.4. Cauchy boundaries

The head-dependency of Cauchy boundaries has noteworthy effects on the handling with eb-POD. While all information regarding Dirichlet and Neumann boundaries is found in the right-hand side of Eq. (16) (in the respective rows of the **b** vector), Cauchy boundaries are different, as explained below on the example of a river hydraulically connected to the aquifer: the flow from a connected river into the underlying cell is calculated (in MODFLOW, for example) as $q_{riv} = C_{riv}(h_{riv} - h)$, where q_{riv} is the flow [L^3/T] into the cell from the river, C_{riv} is the conductance [L^2/T] of the river-bed and h_{riv} is the head in the river [L]. This can be rewritten as $q_{riv} = C_{riv}h_{riv} - C_{riv}h$, resulting in a (groundwater-) head-independent part ($C_{riv}h_{riv}$) that is added to the right-hand side **b** of Eq. (16) (as for Neumann boundaries), and a head-dependent part ($-C_{riv}h$) that is added to the diagonal of the left-hand side **A** of Eq. (16). The eb-POD method adds the original boundary equations to the reduced equation system in Eq. (16), but does so in the reduced space. This means that fluxes through head-dependent boundaries (i.e. Cauchy boundaries) are not computed error-free in the eb-POD model. While the head-independent part stored in **b** is correct and error-free (through its enforcement similar to a Neumann boundary in eb-POD), the head-dependent part in **A** is calculated from the approximated heads ($\mathbf{Ah} \cong \mathbf{APr}$), and is therefore not entirely error-free. This is demonstrated in Section 3.4.

2.3.5. Implications of eb-POD

Looking at the left-hand side matrix of Eq. (15), several things are important. First, the overall size of the eb-POD reduced order model does not change compared to the POD model – both handle full matrices of size $n_r \times n_r$ to find the solution for each time step. Thus, the explicit treatment of boundary conditions in eb-POD comes at no additional computational cost compared to the original POD method. What changed in size, however, are the corresponding projection matrices. While the original POD matrix $\mathbf{P} \in \mathbb{R}^{n_m \times n_r}$ uses the information of n_r singular vectors to span the reduced model subspace, the projection matrix $\mathbf{\Psi} \in \mathbb{R}^{n_m \times n_{r2}}$ retains less information, since $n_{r2} = n_r - n_{bc}$. Practically, this means that the reduced model size n_r has to be chosen larger than the number of independent boundary conditions n_{bc} treated this way. Replacing original POD rows by boundary rows reduces, in theory, the accuracy of the solution within the domain. However, the accurate representation of the boundaries typically counteracts this loss of accuracy, or forces the user to decide on the relative importance of different model aspects. This potential trade-off will be analyzed further in the following sections.

Another potential implementation of the eb-POD method would be the addition of the boundary rows to the original POD projection matrix. Let the original POD matrix be $\mathbf{P}_{orig} \in \mathbb{R}^{n_m \times n_r}$. Now, we create a new, bigger projection matrix $\mathbf{P}_{big} \in \mathbb{R}^{n_m \times n_{big}}$ following the procedure detailed in Section 2.2, but retaining as many additional singular values as we have independent boundaries: $n_{big} = n_r + n_{bc}$. We subject this matrix $\mathbf{P}_{big}$ to the steps detailed in Section 2.3.1, resulting in a new eb-POD projection matrix $\mathbf{\Psi}_{add} \in \mathbb{R}^{n_m \times n_{r2}}$. Note that for this matrix, $n_{r2} = n_{big} - n_{bc}$ and thus $n_{r2} = n_r$. This means there is no loss of information within the domain by going from $\mathbf{P}_{orig}$ to $\mathbf{\Psi}_{add}$. Application of this alternative implementation (not shown) proved that this is indeed the case, and the accuracy of the eb-POD model within the domain is in agreement with the original POD model while retaining the feature of accurate boundary representation. It is important to note, though, that the overall size of the eb-POD ROM is $n_{r2} + n_{bc}$, and thus n_{bc} bigger than the original POD ROM. Therefore, this new eb-POD model is, while more accurate, slower than the original POD model. Thus, the question of subtraction or addition of the n_{bc} boundary condition rows to the ROM is the same trade-off between run time vs. accuracy as the original choice of the reduced model size n_r. For the sake of comparison between POD and eb-POD, it seemed more practical if the reduced model size for both methods is the same, thus fixing run times and exploring the trade-off in accuracy instead of in run time.

2.4. Synthetic model setup

2.4.1. One-dimensional (1D) models

To demonstrate our new method, we use four simple synthetic 1D models with different types of transient boundary conditions as well as a more complex synthetic 3D model. While differing in their boundary conditions, all four 1D models have the same geometry, time settings and aquifer properties. Following earlier POD benchmarking studies (Siade et al., 2010), we set up 1D confined groundwater models that consist of 1 row, 100 columns and 1 layer (a total of 100 model cells). All model cells are 100 m wide (Δx), 1 m deep (Δy) and 100 m high (Δz). The synthetic aquifer's dimensions are therefore 10, $000 \times 1 \times 100$ m. The model simulation time is one year with daily time steps, leading to 365 time steps. Two zones of hydraulic conductivity are used to incorporate some complexity and asymmetry into the model. Two-thirds of the model domain (cells 1–67) have hydraulic conductivity values of 1500 m/d, while the rest of the domain (cells 68–100) has a hydraulic conductivity of 500 m/d. The entire model domain has a specific storage value of 0.005 [1/m].

Fig. 1 shows the different 1D model schematics with the boundary conditions and hydraulic conductivity zones. Table 2 summarizes the differences of the four 1D models and states which boundary conditions are treated explicitly by the eb-POD method. The pumping regime of all models can be seen in Table 1.

The first 1D model, termed M1, serves as a simple proof of concept for the method. The model is set up analogous to the 1D models used in groundwater POD literature (see, for example, McPhee and Yeh, 2008). The left and right model edges are set as constant head boundaries at −1 and −2 m, respectively. The initial groundwater head follows the linear gradient interpolated between these two constant head boundaries. A single pumping well is added to cell 33 ($x = 3250$ m) with a pumping rate of -100 m^3/d which is turned on in the interval $\mathbf{t}_p = [100 - 250]$ days. For this model, only the single Neumann boundary cell (the pumping well) is treated explicitly by eb-POD.

The second 1D model, M2, is more complex. Here we replace the left constant head boundary with an unconnected river (Neumann-type boundary) with a hydrograph taken (with slight adjustments) from a real-world river (see Fig. 4(e)). The initial groundwater head is identical to the remaining constant boundary on the right model edge for the entire domain. Furthermore, a second pumping well is added to the model domain at cell 67 ($x =$ 6650 m) and both wells are pumped with changing pumping rates and partly separate, partly overlapping pumping regimes, ranging from -150 m^3/d to -250 m^3/d. In this model, all three Neumann boundaries (the river and both pumping wells) are treated explicitly by eb-POD.

M3, the third 1D model, is used to show the application to a variable Dirichlet boundary. The model set-up is similar to M1, but the left boundary is changed to a variable head (simulating a seasonal head change with a sine-function with period 1 year and amplitude of ± 1 m). Only this variable Dirichlet boundary is handled explicitly by eb-POD.

The last 1D model, M4, is similar to M2, but connecting the river to the aquifer, which then is a Cauchy-type boundary. As in M2, both river and pumping wells (one Cauchy, two Neumann boundaries) are explicitly treated by eb-POD.

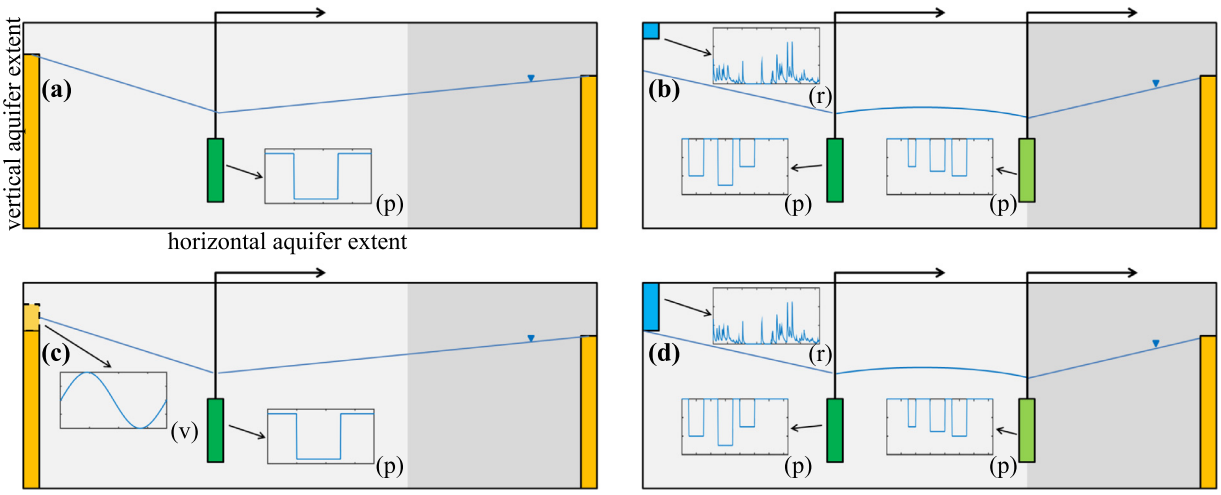

Fig. 1. Schematic view of the four synthetic 1D test models M1 (a), M2 (b), M3 (c) and M4 (d). The gray background delineates the two hydraulic conductivity zones. Yellow bars are (constant or variable) head boundaries, green bars pumping wells and blue bars river boundaries. Input for time-dependent boundaries is shown in the small figures with time on the *x*-axis and pumping rate (*p*), variable head (*v*) or river head (*r*) on the *y*-axis. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Pumping regimes.

	Pumping rate [m^3/d] and pumping interval [d]		
	M1/M3	M2/M4	M5
Pumping well 1	−100 in [100–250]	−200 in [25–75] −250 in [125–175]−150 in [200–250]	−600 in [25–75] −750 in [125–175]−450 in [200–250]
Pumping well 2	–	−150 in [75–100] −175 in [150–200]−200 in [225–275]	−450 in [75–100] −525 in [150–200]−600 in [225–275]

Table 2
Synthetic test model summary.

Model	Left bc	Right bc	Inner bc	bc handled by eb-POD
M1	Const. head (Dirichlet)	Const. head (Dirichlet)	1 × pumping well (Neumann)	1 × pumping well
M2	Uncon. river (Neumann)	Const. head (Dirichlet)	2 × pumping well (Neumann)	River 2 × pumping well
M3	Var. head (Dirichlet)	Const. head (Dirichlet)	1 × pumping well (Neumann)	Var. head
M4	Con. river (Cauchy)	Const. head (Dirichlet)	2 × pumping well (Neumann)	River 2 × pumping well
M5	Uncon. river (Neumann)	Const. head (Dirichlet)	2 × pumping well (Neumann)	River 2 × pumping well

Table 3
3D model M5 hydraulic conductivity values (in m/d).

Zone	K_x (along columns)	K_y (along rows)	K_z (along layers)
(A)	150	50	30
(B)	50	30	20
(C)	80	40	25
(D)	20	15	10

2.4.2. Three-dimensional (3D) model

A fifth synthetic test model, M5, is presented here to show the applicability of the method for a more complex model. M5 is a 3D confined groundwater model with 30 rows, 100 columns and 5 layers, i.e. a total of 15,000 model cells. The size of the model cells is 10 × 10 m horizontally (Δx and Δy, respectively) and 5 m vertically (Δz), resulting in aquifer dimensions of 1000 × 300 × 50 m. Model simulation time is again one year with daily time steps (365). Hydraulic conductivity is assumed to be different for odd- and even-numbered layers, resulting in four zones (labeled A–D) with different hydraulic conductivity values in *x*, *y* and *z* direction (Fig. 2 and Table 3). The specific storage value of the entire model domain is assumed to be 0.005 [1/m].

With respect to the definition of the boundary conditions, model M5 follows the build of M2, with a hydraulically unconnected river (Neumann boundary) at the left model domain (Fig. 2) exhibiting the same hydrograph as used in M2. The right model boundary is a constant head of 0 m. The initial groundwater head is set to 0 m across the entire model domain. Two pumping wells are added in layer 5 (Fig. 2). Pumping rates and intervals for these wells are listed in Table 1. As before, all three Neumann boundaries (one river, two pumping wells) are treated explicitly by eb-POD, as can be seen in Table 2.

2.5. Snapshot selection technique

Snapshot selection is an important factor in POD-based reduced-order modeling and has been investigated in the literature (cf. Siade et al., 2010). The aim of this study, however, is the comparison of standard POD and an extension to that method, the eb-POD. We assume that this comparison is meaningful regardless of the snapshot selection method, provided the applied process is following tested standard procedures and both POD and eb-POD reduced-order models use the same set of snapshots for each model. On that basis, we used a snapshot selection approach mod-

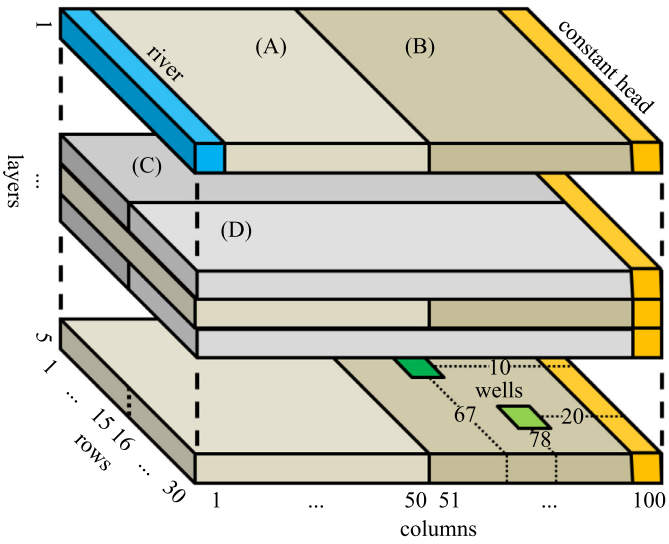

Fig. 2. Schematic view of the 3D test model M5. The colored backgrounds and labels (A)–(D) show the four hydraulic conductivity zones. Yellow blocks are constant head boundaries, green cells pumping wells and blue blocks river boundary. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

ified from Gunzburger et al. (2007) that can be described by the following three steps:

1. identify all drivers of the groundwater system (Dirichlet boundaries, Neumann boundaries (wells, rivers, or recharge, for example), Cauchy-type boundaries) $\rightarrow n_d$
2. compute a steady-state system response to all time-independent boundaries of the groundwater system (constant heads, for example) $\rightarrow \mathbf{h}_{ini}$ and add this steady-state to the snapshot matrix $\rightarrow \mathbf{H}_s = [\mathbf{h}_{ini}]$
3. run the following loop for each driver, $i = 1 \dots n_d$:
   (a) store the current set of snapshots as $\mathbf{H}_{s_{old}} = \mathbf{H}_s$
   (b) set the chosen driver to an arbitrary, but reasonable value constant in time (pumping in the well, river head, etc.)
   (c) "turn off" all other drivers (set pumping to zero, river head to river bottom, variable head to constant, etc.)
   (d) set $\mathbf{h}_0 = \mathbf{h}_{ini}$
   (e) run the model until steady-state is reached, generating system responses (i.e. groundwater heads) at each time step $\mathbf{t} = [0, \dots, n_{td}]$ for this driver $\rightarrow \mathbf{H}_i \in \mathbb{R}^{n_m \times n_{td}}$
   (f) append these responses to the overall snapshot matrix $\rightarrow \mathbf{H}_s = [\mathbf{H}_{s_{old}} \mathbf{H}_i]$

While the original snapshot selection in Gunzburger et al. (2007) extends this approach by running the model with all time-variable drivers "turned on" simultaneously after each step (e) of the loop, our experiments revealed that these additional runs have no influence on the precision or stability of either POD or eb-POD reduced-order models (results not shown). Since they add a significant portion to offline computing time, however, we chose to remove this step from our snapshot selection approach. The chosen approach also follows general guidelines for snapshot selection for POD in groundwater modeling (e.g. Vermeulen et al., 2004a; McPhee and Yeh, 2008). In the literature, one often finds several techniques for snapshot pre-processing, such as mean centering and normalization (cf. Boyce et al., 2015). We could not see any improvement of the POD reduced-order models for our test cases by applying these techniques. In contrary, mean centering of the snapshots before applying the explicit boundary handling for eb-POD resulted in an non-functioning ROM for our test cases. This is because eb-POD's handling of boundary conditions is dependent on the time-variant information contained in the snapshot set that mean centering removes.

The snapshot selection approach described above was used for all different test models. The time for each driver model run to reach steady-state, n_{td}, was chosen to be 100 days. Table 4 summarizes the number of drivers in the system, n_d, resulting snapshot numbers and reduced order model sizes with a chosen retained variance RV of 99.99%. Note that the number of explicitly treated boundaries, n_{bc}, is not the same as the number of drivers in the system, n_d, as only specific boundaries were handled explicitly by eb-POD, as described in Section 2.4. For model M5, $n_{bc} = 3$ (one river and two pumping wells), even though the river spans the whole left-hand side of the first layer (30 cells). Since all river cells are subjected to the same boundary forcing, though, only one single row of the original projection matrix has to be collected for the eb-POD handling of the river.

3. Results

In this section, we apply the eb-POD method to the four 1D models and the 3D model outlined in Section 2.4. We compare original POD and eb-POD in their precision for the original boundary conditions and the general approximation of the groundwater head fields. Note that we will provide run times only for the 3D model, as the 1D models are too simplistic for a meaningful comparison.

3.1. Model M1

Fig. 3(a) shows the head distribution (every ten cells) for all three realizations at two different time steps. One can see a good fit for both ROMs. This can be verified by Fig. 3(b), which shows the absolute maximum head error of both ROMs at each time step. These head errors are very small for most of the simulation, with maximum values of less than 1 cm for the POD and of about 3 cm for the eb-POD model. Table 5 states these errors along with normalized root mean squared errors (NRMSE) in time and space for the groundwater heads for all four models. The biggest errors coincide with the start and stop of pumping at the well, meaning that radical boundary changes are the biggest challenge for the ROMs.

Fig. 3(c) shows the relative flux error at the Neumann boundary (the pumping well) for POD and eb-POD. While the flux error at the start of the pumping period is greater than 5% for the POD model, it is effectively zero for the eb-POD. Note that flux errors for the eb-POD model are not exactly zero due to the accuracy of the applied numerical solvers. Clearly, the eb-POD method successfully integrated the Neumann boundary condition into the ROM. This can also be seen from the low NRMSE values of the Neumann boundary flux at the well shown in Table 6. Flux errors at the boundary only occurred shortly after the start of the pumping phase for this simple model, though.

This improvement at the boundary comes at a cost, though, as was indicated in Section 2.3.5: there is a trade-off between the reduction of the flux error at the boundary and the overall head error for the eb-POD model. In this case, that is true for the magnitude of the head errors at major system disturbances caused by changes in the boundary conditions, while the head errors after some time of continuous pumping are actually smaller for eb-POD than for POD.

3.2. Model M2

Model M2 was built to see whether a more complex setup results in different patterns of POD flux errors, or in more significant changes in eb-POD head errors. Fig. 4(a) shows the head distribution for the two reduced and the full model at different time steps.

Table 4
Reduced-order model summary.

Model	# of drivers n_d	# of snapshots n_s	Reduced model size n_r	# of explicitly treated boundaries n_{bc}
M1	3	300	5	1
M2	4	400	10	3
M3	3	300	8	1
M4	4	400	10	3
M5	4	400	11	3

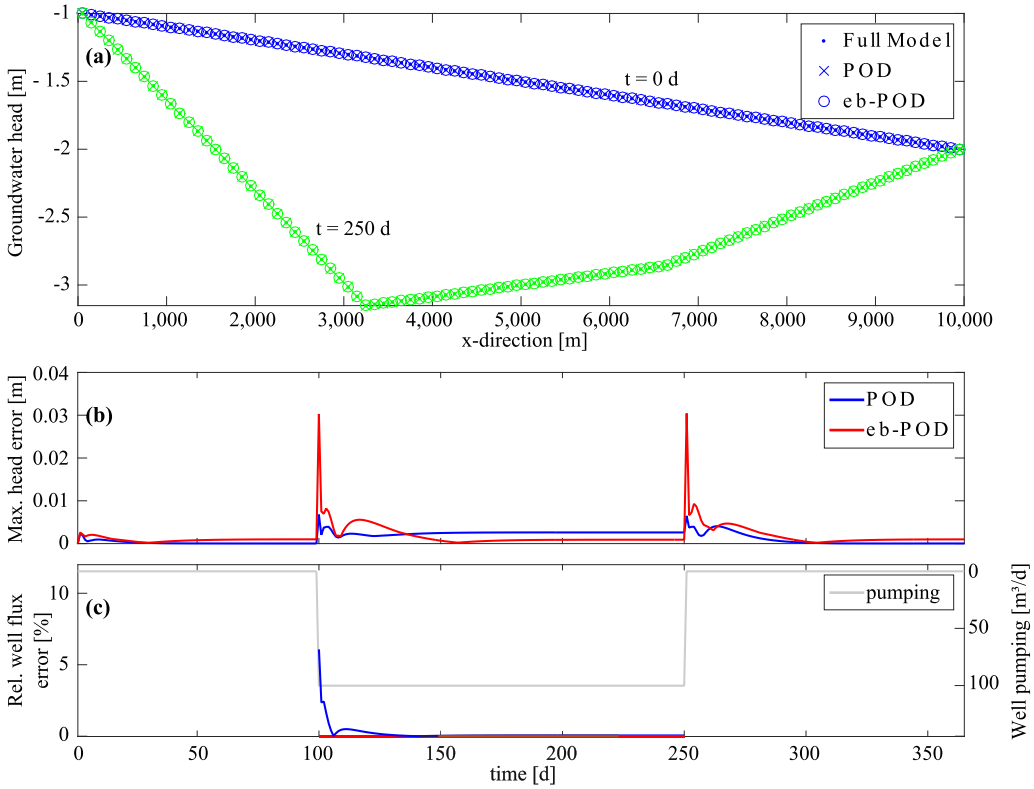

Fig. 3. Model M1: POD and eb-POD for one simple Neumann boundary (well). (a) Water table in the model domain at two different time steps for the full model and both ROMs. (b) Absolute maximum head error over time for both ROMs. (c) Relative well flux error over time along with the pumping rate at the pumping well.

Table 5
Model reduction results for both POD and eb-POD models compared to the full models.

ROM	M1		M2		M3		M4		M5	
	Max error [m]	NRMSE [m]	Max error [m]	NRMSE [m]	Max error [m]	NRMSE [m]	Max error [m]	NRMSE [m]	Max error [m]	NRMSE [m]
POD	0.0068	0.0014	0.0215	0.0010	–	–	0.0178	0.0006	0.0072	0.0008
eb-POD	0.0304	0.0024	0.0904	0.0051	0.0148	0.0048	0.0596	0.0034	0.0098	0.0014

Both reduced models are in good agreement with the full model regarding the water table. The absolute maximum head errors depicted in Fig. 4(b) are about 2 cm for the POD model and peaking to about 9 cm for the eb-POD model. It is easily recognizable that these errors are larger than for model M1.

Fig. 4(c) and (d) show the relative flux errors at the two pumping wells. Flux errors at the wells are up to 5 % for the POD model and negligible for the eb-POD model. While this is very similar to model M1, the patterns of the flux errors are significantly different. The highest errors are at the start of a pumping phase, but they do not diminish in the same manner as in M1.

This can be explained by the POD flux error at the river boundary being highly dynamic, mainly being governed by the input hydrograph (Fig. 4(e)). Comparing the dynamic shapes of the pumping well flux errors with the river flux errors, one can see that the river flux errors strongly influence the pumping well errors, especially for well 2. The eb-POD model treated all three Neumann boundaries accurately, eliminating the flux errors at these boundaries, as is confirmed by low NRMSE values (Table 6). The trade-off compared to maximum head errors is apparent again, though, as the eb-POD has a consistently lower performance in this regard.

3.3. Model M3

We compare only the eb-POD with the full model for model M3, as the standard POD implementation is not applicable for variable Dirichlet boundaries in most cases (as was stated in Section 2.3.2, correct implementation of Dirichlet boundaries in

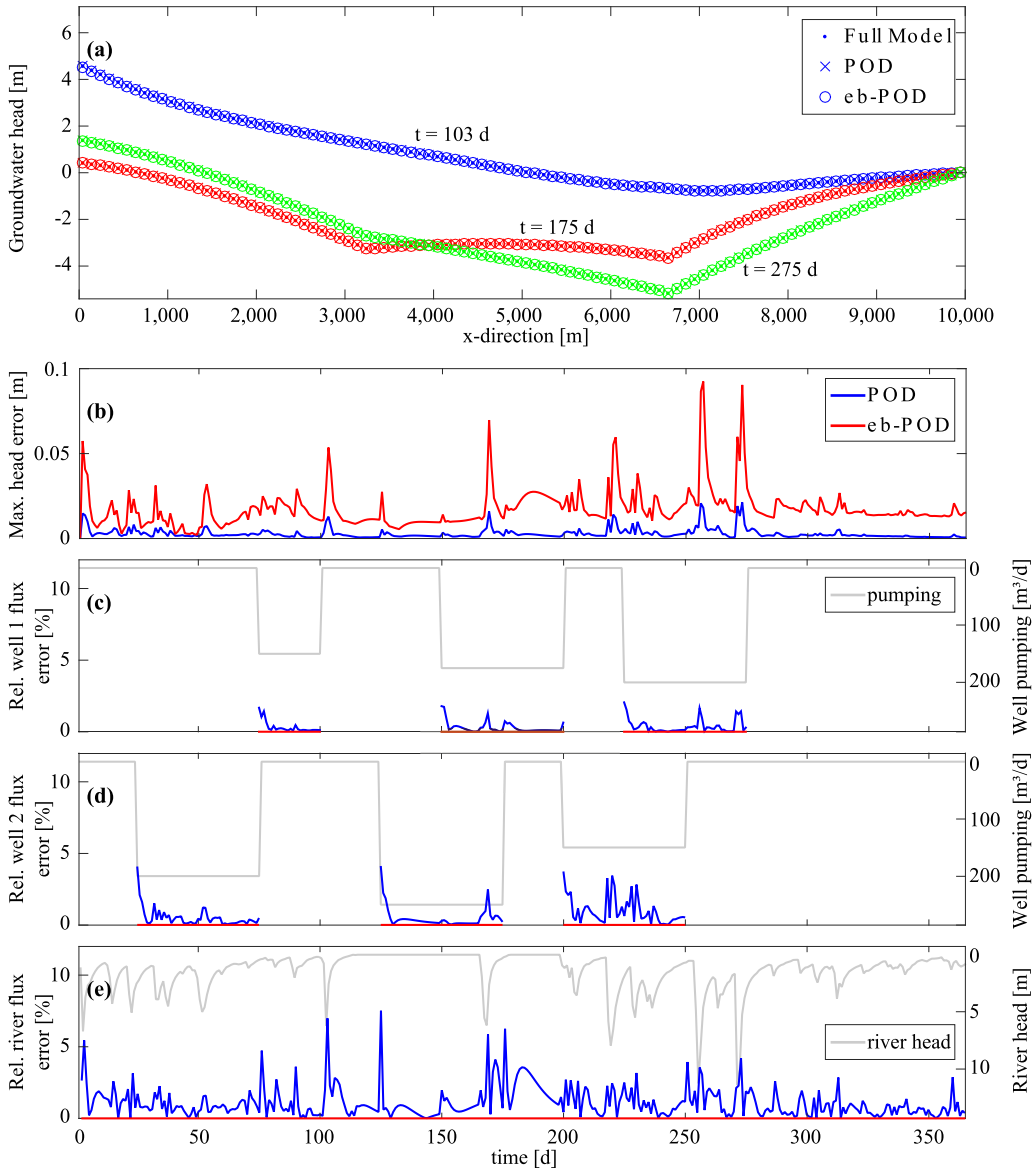

Fig. 4. Model M2: POD and eb-POD for several Neumann boundaries (two wells, one river). (a) Water table in the model domain at three different time steps for the full model and both ROMs. (b) Absolute maximum head error over time for both ROMs. (c), (d) Relative well flux errors over time along with the appropriate pumping rates at both pumping wells. (e) Relative river flux error over time and river head.

Table 6
NRMSE values for both POD and eb-POD models at (eb-POD handled) boundaries.

	M1		M2		M3		M4		M5		
ROM	Well [m³/d]	Well 1 [m³/d]	Well 2 [m³/d]	River [m³/d]	Variable head [m]	Well 1 [m³/d]	Well 2 [m³/d]	River [m³/d]	Well 1 [m³/d]	Well 2 [m³/d]	River [m³/d]
POD	0.0116	0.0203	0.0103	0.0290	–	0.0350	0.0199	0.0227	0.0029	0.0051	0.0085
eb-POD	1.8×10^{-14}	9.3×10^{-15}	9.9×10^{-15}	9.0×10^{-15}	2.5×10^{-12}	2.4×10^{-14}	8.4×10^{-15}	0.0064	5.5×10^{-16}	6.7×10^{-16}	6.7×10^{-5}

standard POD is dependent on several restrictions regarding the ROM). Fig. 5(a) shows the water table and Fig. 5(b) the head error at the variable Dirichlet boundary. Note that the scale of Fig. 5(b) is orders of magnitude smaller than the similar graphs for the other test cases. The eb-POD model reproduces the full model very well, and the errors at the variable Dirichlet boundary are negligible. Absolute maximum head errors of about 1.5 cm are comparable with results from model M1, as is the NRMSE value (cf. Table 5).

3.4. Model M4

The difference between M4 and M2 is the river bed elevation (see Fig. 1). In model M4, the river is connected directly to the aquifer, resulting in a Cauchy-type boundary condition. Head distribution, head errors and pumping well flux errors of the POD and eb-POD models are very similar to model M2 and therefore not shown.

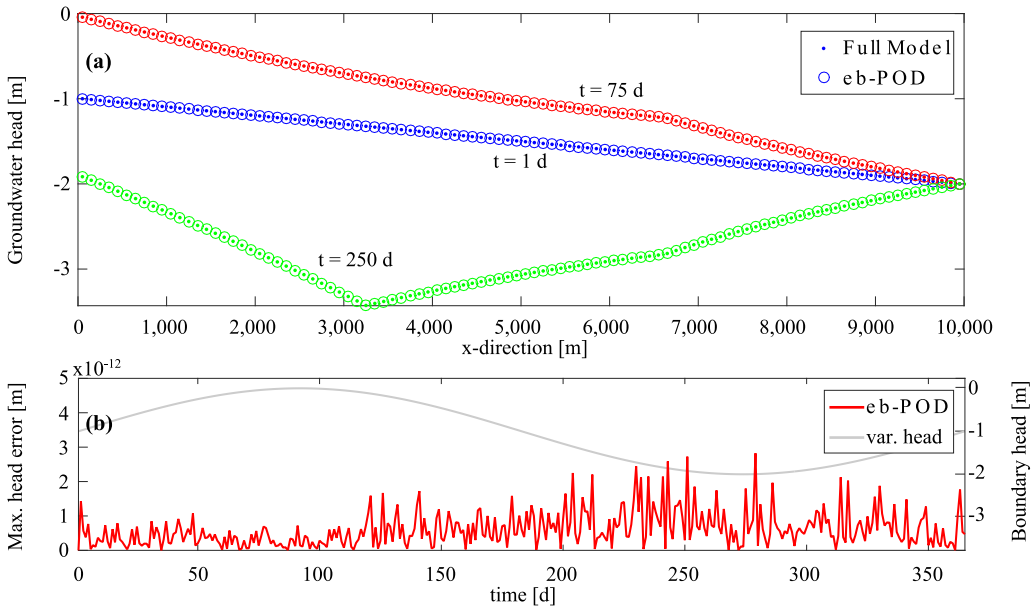

Fig. 5. Model M3: eb-POD for one variable Dirichlet boundary (head). (a) Water table in the model domain at three different time steps for the full model and the eb-POD ROM. (b) Absolute maximum head error over time for eb-POD and change in head at the variable head boundary.

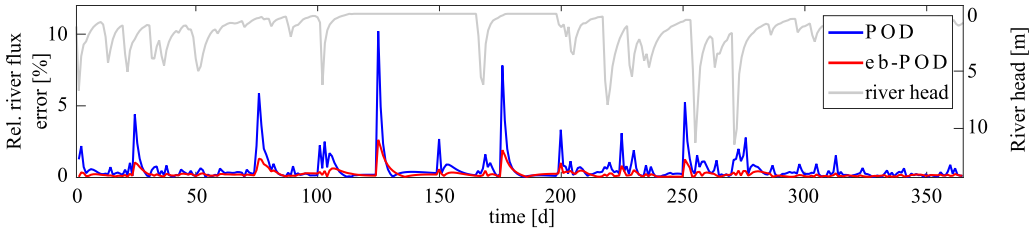

Fig. 6. Model M4: relative flux error over time of POD and eb-POD for Cauchy boundary, along with the river head.

Fig. 6 shows the flux error at the Cauchy boundary. Overall, the flux errors for both ROMs at the Cauchy boundary are larger than they were for the disconnected river boundary (Neumann) in model M2. Flux errors of the eb-POD model at the Cauchy boundary are not zero, contrary to the previous models. This is confirmed by the NRMSE values in Table 6, where the eb-POD model shows significant errors at the Cauchy boundaries, compared to the errors at Neumann boundaries discussed before.

This can be explained by the difference between Neumann and Cauchy boundaries. Neumann boundaries are independent of the groundwater head and therefore only part of the right-hand side vector **b** in Eq. (16). Cauchy boundaries, however, are head-dependent. As was detailed in Section 2.3.4, this means that information regarding their boundary fluxes is stored in both the right-hand side **b** and the left-hand side **A** of Eq. (16). While the first part can be enforced error-free by the eb-POD method, the second part is calculated from the approximated groundwater heads and thus contains errors resulting from the approximation via the subspace projection. This means that the flux error through Cauchy boundaries cannot be eliminated completely, but compared to the original POD model, it leads to a reduction of the error in most cases, as can be seen in Fig. 6.

3.5. Model M5

Model M5 was used for an application of both POD and eb-POD to a larger, more complex 3D model. Fig. 7 shows the groundwater levels in the model domain at two selected times ($t_1 = 175$ d, $t_2 = 225$ d) for the full model, along with errors for the two ROMs as head differences to the heads of the original model. Fig. 7(a) and (b) depicts the depression cones of the two pumping wells in the full model for two times, respectively, along with the different groundwater head values at the left boundary, which are influenced by the river stage. The head errors for both ROMs at the two time steps are very similar in their structure (Fig. 7(c), (e) and (d), (f)). The absolute values are different, though, with slightly larger errors for the eb-POD model. This is in agreement with the results of the other test cases, and is caused by the information loss through the reduction of retained singular values in the projection matrix for eb-POD. This is also confirmed by Fig. 8(a), which shows the maximum head errors of both ROMs over time. Please note that these errors are still small and at most in the range of typical head measurement errors. This is confirmed by head errors and NRMSE values reported in Table 5.

Fig. 8(b) and (c) shows the relative flux errors at the two pumping wells, which are relatively small for the POD model and almost zero for the eb-POD model. The relative flux errors at the river boundary are depicted in Fig. 8(d). Again, the flux errors at this Neumann-type boundary are significant for the POD model, while the errors for the eb-POD model are virtually zero, as confirmed by the NRMSE values in Table 6. Note that for the 3D model, this river boundary spans 30 cells, which could all be represented in the eb-POD matrix in a single line (see Section 2.5). This shows the potential of the eb-POD method for complex groundwater models, since boundary conditions implemented into several cells can be accurately represented in an eb-POD reduced model.

The run times of the 3D models are summarized in Table 7. A run time reduction by a factor of 6 was observed for the on-

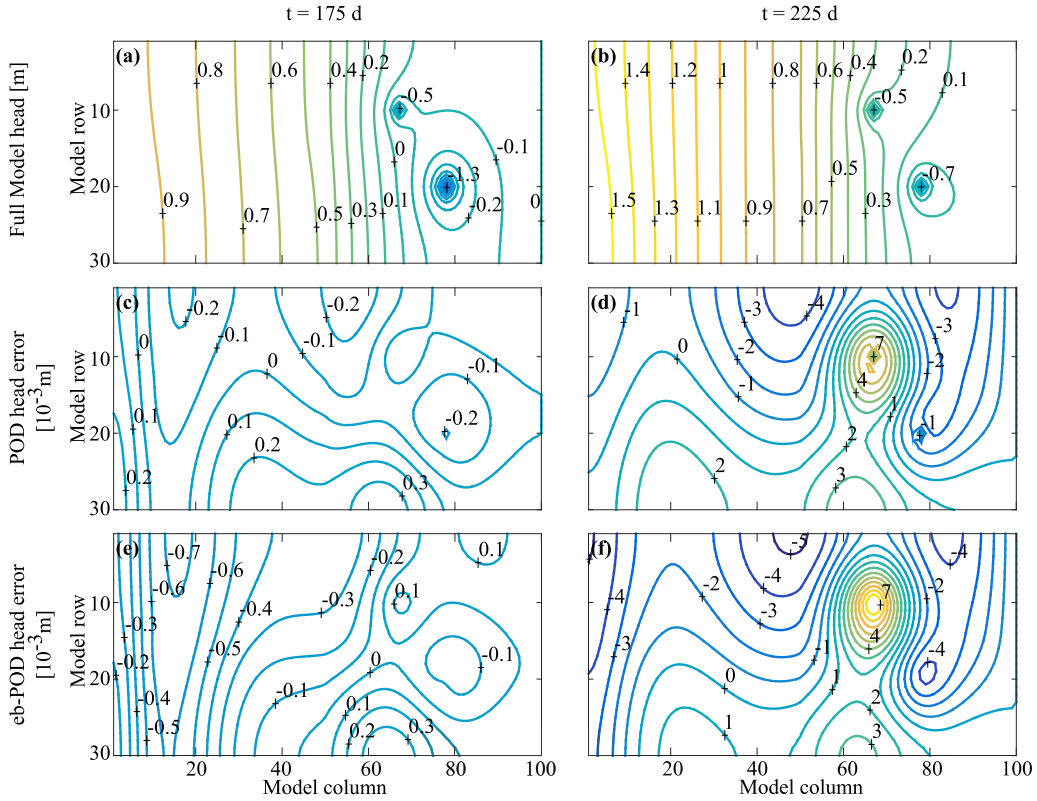

Fig. 7. Model M5: hydraulic head contours [m] for the full model at time steps $t_1 = 175$ d (a) and $t_2 = 225$ d (b), and head field differences [10^{-3} m] of the POD (c and d) and eb-POD (e and f) models.

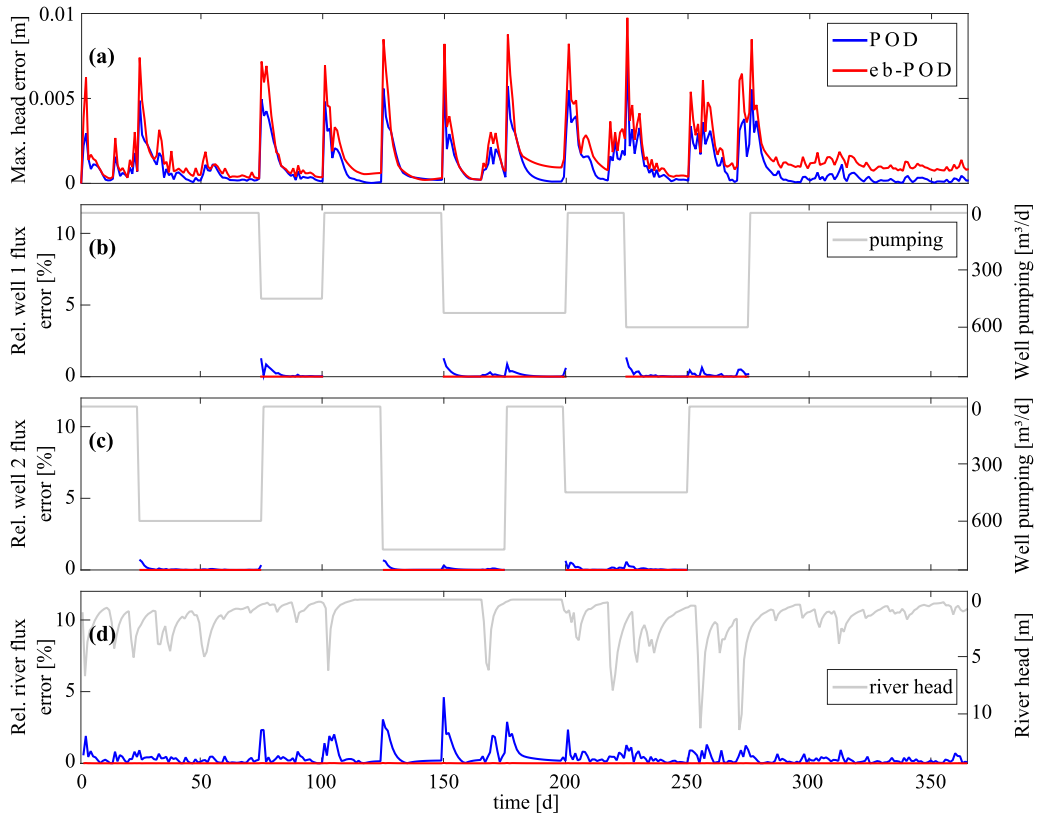

Fig. 8. Model M5: (a) absolute maximum head error over time for both ROMs. (b), (c) Relative well flux errors over time along with the appropriate pumping rates at both pumping wells. (d) Relative river flux error over time and river head.

Table 7
Run time comparison of the full model and both ROMs for the 3D model M5.

	Full model	Snapshots and basis (offline)	POD (online)	eb-POD (online)
Total run time [s]	120	135	20	20
System matrix inversion [s]	0.2	–	0.006	0.006

line portions of the ROMs. This is even higher, more than factor 30, if considering just the timings of the matrix inversion step. Reductions are the same for both ROMs due to their identical sizes. In our test case, the offline portion of calculating the snapshots and creating the projection matrix takes as much time as one full model run. Since it is often not necessary to repeat this step for ROM applications (calculation with different boundaries, for example), this is an acceptable time for pre-processing. Note that run times are taken on a standard desktop-PC for model implementations in Matlab. There is potential to further improve actual run times by, for example, compilation of the code.

4. Discussion

The results presented in Section 3 show that the eb-POD method can be applied for all three types of boundary conditions. Originally, POD models adhere to constant Dirichlet boundaries through pre-processing of snapshots, while variable Dirichlet boundaries could not be implemented this way at all. Alternatively, Dirichlet boundaries can be handled correctly by application of POD to drawdown simulations and calculation of groundwater heads through superposition, but only for linear models. Neumann and Cauchy boundaries are only implicitly handled. This leads to flux errors through these boundaries, which, even in simple 1D models, can become quite significant (>10% relative error). One can also see that changes at certain boundaries influence the flux errors at other boundaries.

The proposed eb-POD method extends the general POD for all three boundary types in a straightforward manner that does not necessitate complicated computations. Model M3 shows the application to a variable Dirichlet boundary, which can only be reproduced by standard POD in special cases. The resulting errors of the eb-POD ROM at the head boundary are negligible. Models M1, M2 and M5 are applications of the method for (different) Neumann boundaries. The explicit treatment of these boundaries through eb-POD leads to a reduction of the flux errors through the boundaries to effectively zero. Model M2 shows that the interplay of several boundaries can lead to substantial miscalculation of these fluxes in POD reduced-order models. Thus, one can assume that justification for the eb-POD method increases with the complexity of the model. Model M5 shows that the method works satisfactory in 3D, allowing the representation of a 30-cell river boundary through one row in the eb-POD projection matrix. The ROMs exhibit a run time reduction by a factor of 6. While model M4 shows that the method can also be applied for Cauchy boundaries, the accompanying flux errors through this boundary type cannot be eliminated completely. Cauchy boundaries are composed of a head-independent (i.e. linear) and a head-dependent (i.e. non-linear) part, where the latter is calculated from the approximated head instead of explicitly enforced in eb-POD, and is therefore error-prone. Nonetheless, the eb-POD method still explicitly adheres to the head-independent part of the Cauchy boundary, thus diminishing the resulting flux error compared to the original POD. All in all, the new eb-POD method successfully handles ROM errors at boundaries.

This elimination (or reduction, for Cauchy-type boundaries) of ROM flux errors through eb-POD comes at a cost: the overall accuracy of the head field in the aquifer is somewhat lower for eb-POD models compared to original POD models of the same build (formed from the same snapshots with the same number of retained singular vectors). As is explained in Section 2.3.5, this is due to the method's handling of the original projection matrix $\mathbf{P}$, which is projected onto a smaller subspace $\mathbf{\Psi}$ of size $n_{r2} = n_r - n_{bc}$ to allow space for the independent boundary condition equations. This projection is accompanied by a loss of information. This also means that the more independent boundary conditions are treated explicitly by eb-POD, the larger the error of the new ROM with respect to the head field. Furthermore, the theoretical maximum of such independent boundaries would be $n_{bc} = n_r - 1$, as a higher number of boundaries would deduct all information of the original projection matrix $\mathbf{P}$. Practically, this limit is even smaller, as a number of n_{bc} very close to n_r would probably result in a highly inaccurate ROM. Note that this limit refers to independent boundaries, though. In groundwater modeling, multiple cells in the model domain are often described by a single boundary. In addition, this only limits the number of explicitly treated boundary conditions, not the overall number of boundaries in the full model.

Essentially, the eb-POD method provides a trade-off for reduced-order modeling with POD: the elimination (Dirichlet and Neumann) or reduction (Cauchy) of boundary errors comes at a cost of increasing inaccuracy of the overall head approximation with increasing number of independent boundaries treated by eb-POD. Fortunately, this trade-off can be easily quantified. Implementation of different boundaries via eb-POD is straight-forward and computationally inexpensive. Therefore, the modeler can test different scenarios regarding boundary condition treatment by eb-POD and compare both reduction of flux errors and increase in head errors with the original POD model and other scenarios. Furthermore, models are often built with a specific purpose in mind and the accuracy may be required at certain locations (and in certain outputs) only. If, for example, the purpose of the model is the quantification of river-aquifer exchange fluxes, the correct calculation of these fluxes is of higher interest than the general head field. In contrast, if head approximations of the reduced-order model at specific measurement points is the main target of the model, one can analyze the distribution of change in head error through application of eb-POD. This targeted analysis allows the modeler an informed decision on which (if any) boundary conditions to treat this way. Similar choices have to be made for general POD reduction as well: the choice of specific relative retained variance RV (see Section 2.2) is one of computational speed versus precision of the reduced-order model. In comparison, the decision regarding boundary condition treatment through eb-POD is one about accuracy at specific boundaries versus accuracy in the overall head field. Fortunately, this is a choice that can be made with quantifiable information on its ramifications, thus handing all power to the user for target-oriented reduced-order modeling.

5. Summary and conclusion

The newly proposed eb-POD method allows the explicit handling of Dirichlet, Neumann and Cauchy boundary conditions for POD-based groundwater model reduction. POD's projection of the system equations onto a subspace does not account for the boundary conditions in the explicit way the original model equations do. By modifying the projection matrix to be orthogonal to the

boundary conditions and then adding the original boundary equations to the final system of equations, the eb-POD method treats the boundary conditions in the same explicit way as the original model.

The method is applied to five synthetic test cases to test it against all three types of boundary conditions. Boundary errors for Dirichlet and Neumann boundaries are successfully eliminated through our new POD extension. Eb-POD reduces flux errors at Cauchy boundaries compared to POD, though not to zero due to their head-dependence. The test cases highlight the new method's ability to severely enhance accuracy of POD ROM's boundary representation even in 3D, while retaining computational time savings of the original POD method.

The implementation of the method results in slightly larger errors in the overall head field, though. This stems from the method's reduction in size of the original POD projection matrix containing the snapshot information. Fortunately, the effect of this trade-off can easily be quantified. We therefore provide a tool for target-orientated decision making on intrinsic boundary accuracy with the eb-POD method. This adds a direct and specific way of controlling the ROM accuracy to the present choice of snapshot selection and retained variance.

While the method has been applied to simple, synthetic test cases in groundwater modeling in this study, the application to and its effects on real-world models is an area of future investigation. Furthermore, the method does not deal with the shortcomings of general POD reduction for non-linear models. It could potentially be combined with techniques like POD-DEIM (see Stanko et al., 2016) to improve its computational performance for non-linear groundwater models.

Acknowledgments

This work was funded by the German Research Foundation (grant WO 1781/1-1). The author would like to thank Michael Sinsbeck at SimTech, Universität Stuttgart for helpful insights and discussions.

References

Boyce, S.E., Nishikawa, T., Yeh, W.W.-G., 2015. Reduced order modeling of the newton formulation of MODFLOW to solve unconfined groundwater flow. Adv. Water Resour. 83, 250–262. https://doi.org/10.1016/j.advwatres.2015.06.005.

Boyce, S.E., Yeh, W.W.-G., 2014. Parameter-independent model reduction of transient groundwater flow models: application to inverse problems. Adv. Water Resour. 69 (0), 168–180. https://doi.org/10.1016/j.advwatres.2014.04.009.

Cosimo, A., Cardona, A., Idelsohn, S., 2016. General treatment of essential boundary conditions in reduced order models for non-linear problems. Adv. Model. Simul. Eng. Sci. 3 (1), 7. https://doi.org/10.1186/s40323-016-0058-8.

Darcy, H., 1856. Les Fontaines Publiques de la Ville de Dijon. Exposition et Application des Principes à Suivre et des Formules à Employer Dans les Questions de Distribution d'eau: Ouvrage Terminé par un Appendice Relatif aux Fournitures d'eau de Plusieurs Villes au Filtrage des Eaux et à la Fabrication des Tuyaux de Fonte, de Plomb, de Tole et de Bitume. Dalmont.

Doherty, J., Christensen, S., 2011. Use of paired simple and complex models to reduce predictive bias and quantify uncertainty. Water Resour. Res. 47 (12), W12534. https://doi.org/10.1029/2011WR010763.

Fienen, M.N., Masterson, J.P., Plant, N.G., Gutierrez, B.T., Thieler, E.R., 2013. Bridging groundwater models and decision support with a Bayesian network. Water Resour. Res. 49 (10), 6459–6473. https://doi.org/10.1002/wrcr.20496.

Galerkin, B.G., 1968. Rods and Plates: Series in Some Questions of Elastic Equilibrium of Rods and Plates. National Technical Information Service.

Gunzburger, M.D., Peterson, J.S., Shadid, J.N., 2007. Reduced-order modeling of time-dependent PDEs with multiple parameters in the boundary data. Comput. Methods Appl. Mech. Eng. 196, 1030–1047. https://doi.org/10.1016/j.cma.2006.08.004.

Harbaugh, A.W., 2005. MODFLOW-2005, The U.S. Geological Survey Modular Ground-Water Model – The Ground-Water Flow Process. Technical Report, 6-A16. U.S. Geological Survey Techniques and Methods.

McPhee, J., Yeh, W.W.-G., 2008. Groundwater management using model reduction via empirical orthogonal functions. J. Water Resour. Plan. Manag. 134 (2), 161–170. https://doi.org/10.1061/(ASCE)0733-9496(2008)134:2(161).

Oladyshkin, S., Nowak, W., 2012. Data-driven uncertainty quantification using the arbitrary polynomial chaos expansion. Reliab. Eng. Syst. Saf. 106, 179–190. https://doi.org/10.1016/j.ress.2012.05.002.

Pinder, G.F., Celia, M.A., 2006. Numerical Solutions of the Groundwater Flow Equation. John Wiley & Sons, Inc., pp. 263–319 https://doi.org/10.1002/0470044209.ch7.

Siade, A.J., Putti, M., Yeh, W.W.-G., 2010. Snapshot selection for groundwater model reduction using proper orthogonal decomposition. Water Resour. Res. 46 (8), W08539. https://doi.org/10.1029/2009WR008792.

Stanko, Z.P., Boyce, S.E., Yeh, W.W.-G., 2016. Nonlinear model reduction of unconfined groundwater flow using POD and DEIM. Adv. Water Resour. 97, 130–143. https://doi.org/10.1016/j.advwatres.2016.09.005.

Taormina, R., Chau, K.-w., Sethi, R., 2012. Artificial neural network simulation of hourly groundwater levels in a coastal aquifer system of the Venice Lagoon. Eng. Appl. Artif. Intell. 25 (8), 1670–1676. https://doi.org/10.1016/j.engappai.2012.02.009.

Ushijima, T.T., Yeh, W.W.-G., 2015. Experimental design for estimating unknown hydraulic conductivity in an aquifer using a genetic algorithm and reduced order model. Adv. Water Resour. 86 (Part A), 193–208. https://doi.org/10.1016/j.advwatres.2015.09.029.

Vermeulen, P., Heemink, A., Stroet, C.T., 2004a. Reduced models for linear groundwater flow models using empirical orthogonal functions. Adv. Water Resour. 27 (1), 57–69. https://doi.org/10.1016/j.advwatres.2003.09.008.

Vermeulen, P.T.M., Heemink, A.W., te Stroet, C.B.M., 2004b. Low-dimensional modelling of numerical groundwater flow. Hydrol. Process. 18 (8), 1487–1504. https://doi.org/10.1002/hyp.1424.

von Gunten, D., Wöhling, T., Haslauer, C., Merchán, D., Causapé, J., Cirpka, O., 2014. Efficient calibration of a distributed PDE-based hydrological model using grid coarsening. J. Hydrol. (Amst.) 519, 3290–3304. https://doi.org/10.1016/j.jhydrol.2014.10.025.

Willcox, K., Megretski, A., 2005. Fourier series for accurate, stable, reduced-order models in large-scale linear applications. SIAM J. Sci. Comput. 26 (3), 944–962. https://doi.org/10.1137/S1064827502418768.

Winton, C., Pettway, J., Kelley, C., Howington, S., Eslinger, O.J., 2011. Application of proper orthogonal decomposition (POD) to inverse problems in saturated groundwater flow. Adv. Water Resour. 34 (12), 1519–1526. https://doi.org/10.1016/j.advwatres.2011.09.007.

Advances in Water Resources 125 (2019) 41–56

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

Simplification error analysis for groundwater predictions with reduced order models

Moritz Gosses[a], Thomas Wöhling[a,b,*]

[a] Department of Hydrology, Technische Universität Dresden, Dresden 01069, Germany
[b] Lincoln Agritech Ltd., Ruakura Research Centre, Hamilton 3240, New Zealand

ARTICLE INFO

Keywords:
Model reduction
Paired model analysis
Predictive simplification error
Regional groundwater modeling

ABSTRACT

Groundwater resource management often requires detailed numerical models that make calibration and predictive uncertainty analysis computationally challenging. Reduced order models (ROMs) alleviate the computational burden but can potentially lead to predictive bias and underestimation of uncertainty. A paired model approach has previously been proposed to estimate the predictive uncertainty of models compared to highly complex, synthetic realities. This approach is modified to analyze and compare the simplification error for groundwater predictions of a real-world numerical MODFLOW model of the Wairau Plains Aquifer in the Marlborough Region of New Zealand. Two different ROM types were applied in this study to predict groundwater heads, spring discharge and river–groundwater exchange fluxes: (1) a drastically simplified MODFLOW model, and (2) artificial neural networks (ANNs). The different ROMs exhibit very different patterns of bias and magnitude of model simplification error. The method accurately captures the simplification error for most predictions by the MODFLOW model, but underestimates the error for predictions highly dependent on the variability of the complex model. The simplified MODFLOW model shows significant parameter surrogacy and non-optimality of simplification, thus questioning the adherence to expert-knowledge based parameter limits. For predictions where historic data sets are available, ANNs provide superior predictive power. However, they cannot be applied to predictions of data types and locations not contained in the calibration data set. For those predictions, simplified numerical models can be applied with varying degree of accuracy.

1. Introduction

Modern groundwater management is often supported by the application of groundwater models. These models integrate available system knowledge (e.g. measurements of system states and expert information) and are used to predict system states under different (future) settings. Differences in modeling approaches are accompanied by a large diversity in model complexity, input data size and model run time, among other factors.

Typically, complex physically-based models are required to represent the spatial variability of coupled groundwater flow processes and to incorporate existing knowledge. Unfortunately, with increasing complexity these models become arduous to handle due to large run times and problems of numerical stability. This vastly restricts the model's applicability for model calibration and uncertainty analysis. A common solution to this problem is model simplification. Different approaches to generate reduced-order models (ROMs) have been developed and applied in the past decades. Three categories of model reduction are most common in groundwater modeling: data-driven methods, projection-based methods and upscaling methods. Data-driven methods such as Bayesian networks (Fienen et al., 2013), artificial neural networks (Taormina et al., 2012) or some types of polynomial chaos expansion (Oladyshkin and Nowak, 2012) result in very fast simple models that reproduce data time series relying only on few input data sources. Due to their "black-box" character, they are limited in their applicability to data types and locations contained in the training data set and do not incorporate any system knowledge or underlying physical principles. Projection-based methods such as proper orthogonal decomposition (POD) (e.g. Boyce et al., 2015; Siade et al., 2012; Ushijima and Yeh, 2015) retain the physical properties of the full model while speeding up run-times by solving the model equations in a subspace computed off-line from full model simulations. Recent studies tried to tackle the method's limitations regarding model non-linearity (Stanko et al., 2016) and boundary representation (Gosses et al., 2018), but applications of POD are still hampered by its complexity in setup, which usually necessitates altering source code of the applied modeling tools to access the model equation systems. Lastly, upscaling methods (e.g. Dagan et al., 2013; von Gunten et al., 2014; Wood, 2009) employ grid coarsening and/or homogenized parameterization to reduce computational times, but are limited by a lack of system detail. In conclusion, model simplifi-

* Corresponding author at: Department of Hydrology, Technische Universität Dresden, Dresden 01069, Germany.
E-mail address: thomas.woehling@tu-dresden.de (T. Wöhling).

https://doi.org/10.1016/j.advwatres.2019.01.006
Received 6 July 2018; Received in revised form 11 January 2019; Accepted 11 January 2019
Available online 15 January 2019

cation tackles the issues of long run times and numerical instability with a variety of different methods which all come at some costs.

A possible solution to the complementary problems of simple and complex models is a paired model approach where the ROM is calibrated against outputs of the complex model. Lødøen and Tjelmeland (2010) and Scheidt et al. (2011) applied such methods in the field of reservoir modeling for petroleum. In groundwater, Doherty and Christensen (2011) and Watson et al. (2013) developed a paired model approach to reduce predictive bias and for uncertainty analysis. They analyzed performance and shortcomings of synthetic groundwater models in comparison to virtual realities. They successfully computed predictive bias and uncertainty and gained insights into causes of uncertainty and the shortcomings of model parameterizations. However, they applied their paired model analysis only to synthetic models as a proof of concept. Furthermore, the simplified models used in their studies are mainly simplifications of the complex model in the computational grid and the model parameters, thus considering only a small portion of possible ROMs in their analysis.

In this study, we apply the methodology developed by Doherty and Christensen (2011) and Watson et al. (2013) to an existing complex model of a real-world aquifer instead of a virtual reality. This distinction allows us to assess the predictive simplification error of the simple model and test it against (a) the simulation results of the calibrated complex model and (b) real-world data pertaining to some of the model predictions. Furthermore, we use two distinctly different ROMs in the study: a simplified physically-based MODFLOW model, and a collection of ANNs as application of very fast, data-driven "black box" models.

This paper is structured as follows. In Section 2, we summarize the underlying subspace theory of the paired model approach. Here, we also describe the extension of the methodology to test the ROMs against the calibrated complex model and real-world data. Section 3 outlines the case study of the Wairau Plain Aquifer, the complex model and both ROM types, as well as the predictions used in the paired model analysis. Section 4 presents the results of the study. The findings are discussed and conclusions are drawn in Section 5.

2. Concepts and theory

This study utilizes a paired simple–complex model approach first presented by Doherty and Christensen (2011) and Watson et al. (2013) to study the simplification error associated with using ROMs for predictions in real-world groundwater system. Since we use this method in a slightly different context and extend it to real-world data, we will first outline the differences to the original method here. The main assumptions and differences between the original approach and our methodology is summarized in Fig. 1.

Doherty and Christensen (2011) and Watson et al. (2013) postulate that their complex models are equal to reality. Outputs of these complex models are deemed to be free of bias and/or error. This allows them to analyze the total predictive uncertainty of their simple model predictions with the paired model approach. Furthermore, they can treat actual field data sets as another potential complex model output and use it for bias correction.

In contrast, the complex model presented in this paper is considered to be an imperfect representation of reality (as all models are by definition). While it is built to accommodate all available system knowledge and calibrated thoroughly, its outputs are subject to (mostly unknown) bias and/or error. The paired model approach presented here can only analyze bias and error between the complex and simple models. This is therefore not the total predictive uncertainty, but will be called simplification error instead. Due to this definition, the bias-correction has to be undertaken on basis of the calibrated complex model output instead of actual measurements. Differing conceptually from the original approach, real-world observations can be used here to estimate the prediction-specific error between complex model and reality.

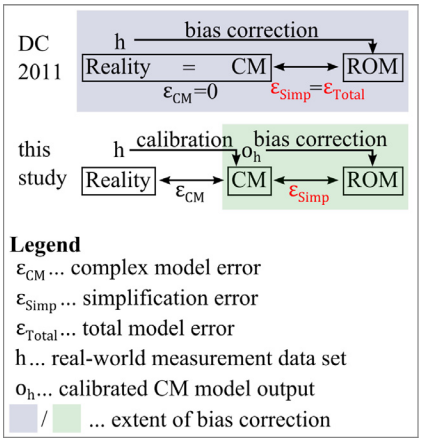

Fig. 1. Comparison of the proposed method in this study with the original method by Doherty and Christensen (2011) (labeled DC 2011).

The underlying concepts of the paired model analysis are based on the assumptions of model linearization. The non-linearity present in most complex groundwater models effects the results of the analysis, but poses no restriction to the applicability of the method presented in the following.

2.1. Linear models: solution space and null-space

Let the following equation denote a linear model:

$$\mathbf{h} = \mathbf{Z}\mathbf{k}+\boldsymbol{\varepsilon}, \tag{1}$$

where the vector $\mathbf{h}$ represents the calibration data set, $\mathbf{k}$ is a vector of model parameters, the matrix $\mathbf{Z}$ denotes the action of the model under calibration condition and the vector $\boldsymbol{\varepsilon}$ describes (random) measurement noise. We define the vector $\underline{\mathbf{k}}$ as the set of model parameters that produce a sufficiently calibrated model. In linear algebra, a matrix $\mathbf{Z}$ has a null-space if the following equation is true for a non-zero vector $\mathbf{k}_n$:

$$\mathbf{0} = \mathbf{Z}\mathbf{k}_n. \tag{2}$$

This is usually true for every linearized form of a complex spatially explicit groundwater model. Combining Eqs. (1) and (2) and introducing a calibrated parameter set $\underline{\mathbf{k}}$ gives

$$\mathbf{h} = \mathbf{Z}(\underline{\mathbf{k}} + \mathbf{k}_n) + \boldsymbol{\varepsilon}. \tag{3}$$

Since there are typically multiple non-zero vectors $\mathbf{k}_n$, Eq. (3) illustrates the problem of non-uniqueness in solving the inverse model problem that is well known to groundwater modelers. For the concepts described in the following, it is important to note that the solution space component of the parameter vector ($\underline{\mathbf{k}}$) can be inferred from the calibration data set $\mathbf{h}$ through the model, while the null-space component ($\mathbf{k}_n$) can not.

2.2. Model simplification: concept and calibration

In this section, model simplification is described by an omission of model parameters from the complex model. Note that in this definition, parameter omission can also represent omission of processes in the simple model. This leads to the definition of

$$\mathbf{k} = \begin{bmatrix}\mathbf{k}_s \\ \mathbf{k}_o\end{bmatrix}, \tag{4}$$

where $\mathbf{k}_s$ describe the simple model parameters and $\mathbf{k}_o$ are the omitted parameters. Combining Eq. (4) with Eq. (1) gives

$$\mathbf{h} = \mathbf{Z}\mathbf{k} + \boldsymbol{\varepsilon} = \begin{bmatrix}\mathbf{Z}_s & \mathbf{Z}_o\end{bmatrix}\begin{bmatrix}\mathbf{k}_s \\ \mathbf{k}_o\end{bmatrix} + \boldsymbol{\varepsilon} = \mathbf{Z}_s\mathbf{k}_s + \mathbf{Z}_o\mathbf{k}_o + \boldsymbol{\varepsilon}, \tag{5}$$

where $\mathbf{Z}_s$ and $\mathbf{Z}_o$ denote portions of the system matrix $\mathbf{Z}$ pertaining to $\mathbf{k}_s$ and $\mathbf{k}_o$, respectively. We now assume that Eq. (2) also holds for $\mathbf{Z}_s$, i.e. that the parameter solution of the simple model is also non-unique. For the methodology presented here, it is assumed that the parameter estimation for $\mathbf{k}_s$ is performed using principal component regression (PCR (Mandel, 1982), often referred to simply as singular value decomposition, SVD, in groundwater literature). This does not necessarily have to be the chosen method for solving the inverse problem (and was not used for the model presented here), but is useful here to understand the concepts described below. Singular value decomposition of $\mathbf{Z}_s$ is written as:

$$\mathbf{Z}_s = \mathbf{USV}^t, \tag{6}$$

where $\mathbf{S}$ holds the singular values of $\mathbf{Z}_s$ and $\mathbf{U}$ and $\mathbf{V}$ the left- and right-singular vectors (which are orthonormal eigenvectors of $\mathbf{Z}_s\mathbf{Z}_s^t$ or $\mathbf{Z}_s^t\mathbf{Z}_s$), respectively. $\mathbf{V}$, $\mathbf{S}$ and $\mathbf{U}$ are partitioned into their solution space ($\mathbf{V}_1$, $\mathbf{S}_1$ and $\mathbf{U}_1$) and null-space ($\mathbf{V}_2$, $\mathbf{S}_2$ and $\mathbf{U}_2$) counterparts (c.f. Aster et al. (2011) for details). Thus, the calibrated simple model parameter set $\underline{\mathbf{k}}_s$ is given as:

$$\underline{\mathbf{k}}_s = \mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^t\mathbf{h} - \mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^t\boldsymbol{\varepsilon}. \tag{7}$$

From Eq. (7), the calibrated complex model parameter set is written as:

$$\underline{\mathbf{k}} = \begin{bmatrix}\underline{\mathbf{k}}_s \\ \mathbf{k}_o\end{bmatrix} = \begin{bmatrix}\mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^t \\ \mathbf{0}\end{bmatrix}\mathbf{h} - \begin{bmatrix}\mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^t \\ \mathbf{0}\end{bmatrix}\boldsymbol{\varepsilon}. \tag{8}$$

Note that the $\mathbf{0}$'s on the right-hand side do not imply that $\mathbf{k}_o = \mathbf{0}$, but only that calibration of the simple model can not be used to infer the omitted parameters $\mathbf{k}_o$. Combining Eq. (8) with Eq. (5) leads to:

$$\underline{\mathbf{k}} = \begin{bmatrix}\underline{\mathbf{k}}_s \\ \mathbf{k}_o\end{bmatrix} = \begin{bmatrix}\mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^t \\ \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{Z}_s & \mathbf{Z}_o\end{bmatrix}\begin{bmatrix}\mathbf{k}_s \\ \mathbf{k}_o\end{bmatrix} - \begin{bmatrix}\mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^t \\ \mathbf{0}\end{bmatrix}\boldsymbol{\varepsilon}. \tag{9}$$

Eq. (9) (Eq. (11) in Doherty and Christensen, 2011) shows the definition of the calibrated complex model parameter set $\underline{\mathbf{k}}$ in regard to the complex model solution space ($\mathbf{V}_1$, $\mathbf{S}_1$ and $\mathbf{U}_1$) and the partitioned system matrix $\begin{bmatrix}\mathbf{Z}_s & \mathbf{Z}_o\end{bmatrix}$. It is used as a basis for the analysis between simple and complex model parameters in the following Section 2.3.

2.3. The relationship between simple and complex model parameters

It is desirable that much of the parameter detail of the complex model that can be explained by the data is retained in the parameter set of the simple model. As it is typically not the case that the parameters themselves are kept in the simple model (because often model simplification means parameter number reduction), we need to analyze how much of the parameter detail is retained in the simple model. To analyze the relationship between simple and complex model parameters, we disregard measurement noise. Note that this is done only for ease of understanding, and that the presence of measurement noise does not invalidate the following analysis. Eq. (9) is rewritten to solve for the calibrated simple model parameter set $\underline{\mathbf{k}}_s$ as

$$\underline{\mathbf{k}}_s = \mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^t\left[\mathbf{Z}_s\mathbf{k}_s + \mathbf{Z}_o\mathbf{k}_o\right]. \tag{10}$$

Combining Eq. (10) with Eq. (6) results in:

$$\underline{\mathbf{k}}_s = \mathbf{V}_1\mathbf{V}_1^t\mathbf{k}_s + \mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^t\mathbf{Z}_o\mathbf{k}_o. \tag{11}$$

From Eq. (11), it becomes obvious that a calibrated simple parameter is usually composed of two parts: the first term on the right-hand side of Eq. (11) is the projection of the simple model parameter set $\mathbf{k}_s$ onto its solution space, while the second term is dependent on the omitted complex parameter set $\mathbf{k}_o$. The second term is called "parameter surrogacy", where the simple model parameter is not only made up by a physical property (a hydraulic conductivity zone value, for example), represented by the first term, but also can represent certain aspects of omitted complex parameters (hydraulic conductivity from another area, or storage values, for example). To analyze this further, we adapt Eq. (22) from Watson et al. (2013) to our definition of an imperfect complex model. Let us start with the following equation that denotes the fitting of the simple model parameters $\mathbf{k}_s$ against the output of the complex model $\mathbf{o}$ by linear regression (equivalent to the first term of the right-hand side of Eq. (22) in Watson et al., 2013):

$$\mathbf{k}_s = (\mathbf{Z}_s{}^t\mathbf{Z}_s)^{-1}\mathbf{Z}_s{}^t\mathbf{o}. \tag{12}$$

We can substitute the complex model by adapting Eq. (1) (substituting $\mathbf{h}$ by $\mathbf{o}$) and ignoring measurement noise into Eq. (12) to get (the second term of the right-hand side of Eq. (22) in Watson et al., 2013):

$$\mathbf{k}_s = (\mathbf{Z}_s{}^t\mathbf{Z}_s)^{-1}\mathbf{Z}_s{}^t\mathbf{Z}\mathbf{k}. \tag{13}$$

By defining $\mathbf{k}_s = \mathbf{L}\mathbf{k}$ (Eq. (20) in Watson et al., 2013) and Eq. (13) we arrive at:

$$\mathbf{L} = (\mathbf{Z}_s{}^t\mathbf{Z}_s)^{-1}\mathbf{Z}_s{}^t\mathbf{Z}. \tag{14}$$

This $\mathbf{L}$ matrix relates the simple model parameter space $\mathbf{k}_s$ to the complex model parameter space $\mathbf{k}$. It allows the investigation of the influence of different complex model parameters onto simple model parameters, and is thus a way to reveal the influence of parameter surrogacy contained in the simple model parameters.

2.4. Components of simplification error for ROM predictions

Making predictions of future and/or unknown system states is a crucial part in environmental modeling, and concomitant with this is the analysis of uncertainty associated to predictions. A prediction is denoted in the linear model approach as

$$s = \mathbf{y}^t\mathbf{k}, \tag{15}$$

where s is a scalar describing the prediction, and the vector $\mathbf{y}$ defines the sensitivities of the prediction to the model parameters $\mathbf{k}$. Separating this equation into simple and omitted parameter components, we get

$$s = \mathbf{y}_s^t\mathbf{k}_s + \mathbf{y}_o^t\mathbf{k}_o. \tag{16}$$

If we now postulate that a prediction made by the calibrated simple model, $\underline{s}$, is written as

$$\underline{s} = \mathbf{y}_s^t\underline{\mathbf{k}}_s, \tag{17}$$

then the error of the simple model prediction in comparison to the complex model prediction is defined as

$$\underline{s} - s = \mathbf{y}_s^t\underline{\mathbf{k}}_s - \mathbf{y}_s^t\mathbf{k}_s - \mathbf{y}_o^t\mathbf{k}_o. \tag{18}$$

Combining Eq. (18) with Eq. (7) (while again ignoring measurement noise $\boldsymbol{\varepsilon}$ for clarity) we get:

$$\underline{s} - s = \mathbf{y}_s^t\mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^t\mathbf{h} - \mathbf{y}_s^t\mathbf{k}_s - \mathbf{y}_o^t\mathbf{k}_o. \tag{19}$$

We can now substitute Eq. (5) into Eq. (19) and simplify utilizing to the orthonormality of $\mathbf{V}$ to get to:

$$\underline{s} - s = -\mathbf{y}_s^t(\mathbf{I} - \mathbf{V}_1\mathbf{V}_1^t)\mathbf{k}_s + \mathbf{y}_s^t\mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^t\mathbf{Z}_o\mathbf{k}_o + \mathbf{y}_s^t\mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^t\boldsymbol{\varepsilon} - \mathbf{y}_o^t\mathbf{k}_o. \tag{20}$$

Finally, we can rearrange Eq. (20) and utilize the relationship $\mathbf{V}_1\mathbf{V}_1^t + \mathbf{V}_2\mathbf{V}_2^t = \mathbf{I}$ (Aster et al., 2011) to describe the simplification error of the model prediction (see Doherty and Christensen (2011) for more details):

$$s = \underline{s} + \mathbf{y}_s^t\mathbf{V}_2\mathbf{V}_2^t\mathbf{k}_s + \mathbf{y}_s^t\mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^t\boldsymbol{\varepsilon} - \mathbf{y}_s^t\mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^t\mathbf{Z}_o\mathbf{k}_o + \mathbf{y}_o^t\mathbf{k}_o. \tag{21}$$

term 0 term 1 term 2 term 3 term 4

In other words, Eq. (21) (Eq. (36) in Doherty and Christensen (2011)) defines the complex model prediction by the composition of the prediction of the calibrated simple model (term 0) and four terms (terms 1–4) describing the difference between the two predictions. The different contributions to the simplification error of the prediction and their implications are best evaluated by analyzing a s versus $\underline{s}$ scatter plot, again summarizing Doherty and Christensen (2011).

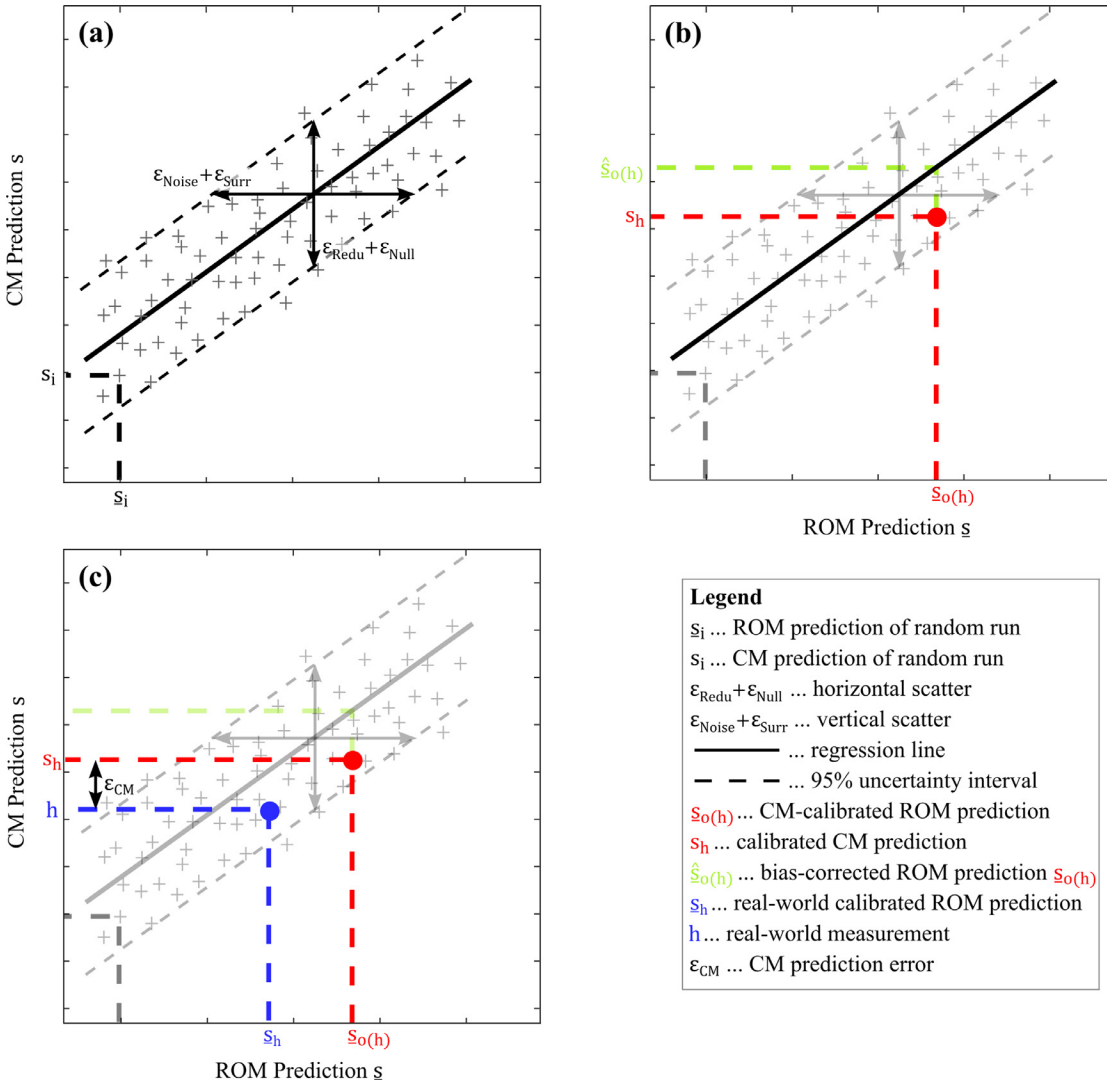

Fig. 2. Schematic of a paired model analysis scatter plot. (a) shows the scatter points, the regression line, its uncertainty intervals and the horizontal and vertical scatter contributions. (b) adds the bias correction of the (CM calibrated) ROM prediction. (c) shows the CM prediction error.

Assume that a number of complex model realizations, $i = 1, \ldots, n$, were randomly generated, each resulting in a set of complex model parameters $\mathbf{k}_i$. The complex model is run (not calibrated) one time with each of these parameter sets to generate a corresponding complex model output vector $\mathbf{o}_i$. To this output $\mathbf{o}_i$, random measurement error $\boldsymbol{\epsilon}_i$ is added to incorporate the influence of term 2 of Eq. (21). Now, the simple model is calibrated against each of the n complex model outputs $\mathbf{o}_i + \boldsymbol{\epsilon}_i$ to derive a simple model parameter set $\underline{\mathbf{k}}_{s,i}$ associated with each complex model parameter set $\mathbf{k}_i$. For each of these n parameter pairs, we compute the prediction of interest, thus obtaining the complex model prediction s_i and the simple model prediction $\underline{s}_i$. As this was done for n model pairs, we then compose a scatter plot of s versus $\underline{s}$ as schematically depicted in Fig. 2(a). This is the basis for the paired model analysis described in Section 2.5.1.

To better associate the different features of these scatter plots with the different terms of Eq. (21), their influence on the shape of the scatter plot is analyzed individually. Again, a detailed description can be found in Doherty and Christensen (2011), and is only partially repeated here to provide the context necessary for the evaluation of the study's results.

If the simple model is a perfect representation of the complex model, Eq. (21) would simplify to $s = \underline{s}$, thus yielding a scatter plot of points perfectly forming a line of slope 1 with intercept 0. Therefore, the different contributions of model simplification error result in a deviation of points from this line.

Term 1 of Eq. (21), $\mathbf{y}_s^t \mathbf{V}_2 \mathbf{V}_2^t \mathbf{k}_s$, denotes the influence of the null space onto the simplification error. The null space (of the simple model) does not affect the simple model prediction $\underline{s}$ (by definition), thus causing the vertical scatter in the plot (due to its influence on the complex model prediction s). Term 2 of Eq. (21), $\mathbf{y}_s^t \mathbf{V}_1 \mathbf{S}_1^{-1} \mathbf{U}_1^t \boldsymbol{\epsilon}$, constitutes the influence of the measurement error $\boldsymbol{\epsilon}$. As the measurement error only pertains to the simple model (calibrated against complex model outputs with the addition of random but predefined measurement error), this leads to horizontal scatter (affecting $\underline{s}$ and not s). Term 3 of Eq. (21), $\mathbf{y}_s^t \mathbf{V}_1 \mathbf{S}_1^{-1} \mathbf{U}_1^t \mathbf{Z}_o \mathbf{k}_o$, is the representation of parameter surrogacy (see Eq. (11) and its explanation). This contribution of parameters and processes that are not represented in the simple model parameters only affects the simple model prediction $\underline{s}$ (as the complex model prediction s depends on the full parameter set), therefore causing horizontal scatter. Finally, term 4 of Eq. (21), $\mathbf{y}_o^t \mathbf{k}_o$, is the influence of the omitted complex model parameter detail on the predictive error. Thus, it only acts upon s and not $\underline{s}$, which results in vertical scatter.

In summary, a perfect simple model would result in s_i, $\underline{s}_i$ points perfectly sitting on a line with slope 1 and intercept 0. Measurement error in calibration of the simple model and the effects of parameter surrogacy

lead to horizontal scatter around this line. Furthermore, vertical scatter can originate from null space contribution and the influence of the omitted parameters and processes. Note that the extent of vertical scatter denotes the simplification error of the model prediction here, while in Doherty and Christensen (2011), it signifies the (total) postcalibration uncertainty of the simple model prediction because they assume that the complex model is reality. Further discussion of the pattern of the scatter plots can be found in Section 2.5.2 as well as in Sections 4 and 5.

2.5. Paired simple-complex model analysis

2.5.1. Methodology

In this section, we briefly summarize the methodology for the paired model analysis that applies the theoretical concepts detailed in Section 2.4 in this study. The methodology was first presented in Doherty and Christensen (2011) and is extended here to accommodate the changes in methodology presented at the beginning of Section 2. The individual steps of the methodology are as follows:

1. Create a complex model $\mathbf{Z}$ with the parameter set $\mathbf{k}$ of the system of interest, incorporating all available system knowledge in process representation and parameterization.
2. Construct n different model parameter sets $\mathbf{k}_i$ (for $i = 1, \dots, n$) and run the model for each of these parameter realizations. These complex model parameter realizations $\mathbf{k}_i$ should represent the unknown variability of the different parameters and thus be random in nature, but the statistics (such as type of distribution, mean and standard deviation etc.) should be informed by the data or expert knowledge about the system. For each of these n random model runs, the output $\mathbf{o}_i$ is recorded.
3. Build a simple model $\mathbf{Z}_s$ that represents a simplified version of the system under study and runs fast enough to allow stochastic simulation.
4. Calibrate the simple model $\mathbf{Z}_s$ against each of the outputs of the complex model $\mathbf{Z}$ corrupted with appropriate measurement errors, $\mathbf{o}_i + \epsilon_i$. Thus, a calibrated simple model realization is obtained for each random complex model realization.
5. Calculate the individual model predictions $\mathbf{s}_i^j$ and $\underline{s}_i^j$ for both the complex and the simple model realizations and all $j = 1, \dots, m$ different predictions.
6. Create a scatter plot of simple model predictions $\underline{s}_i^j$ versus complex model predictions s_i^j for each prediction j. A linear regression line $s = a\underline{s} + b$ is fitted to the $[s_i; \underline{s}_i]$ pairs for each jth prediction and added to the plot. Obviously, any other function could be used instead of a linear regression without loss of generality. Furthermore, standard deviations σ of the regression are calculated and the corresponding 95% confidence intervals ($\pm 2\sigma$, assuming normal distribution) are added to the scatter plot. The confidence intervals are used to discern the contributions to predictive uncertainty by the four terms of Eq. (21). This is shown in Fig. 2(a)and is discussed in detail in Section 2.5.2.
7. Calibrate the simple model $\mathbf{Z}_s$ using the real-world measurement data set $\mathbf{h}$ and compute the corresponding model predictions $\underline{\mathbf{s}}_\mathbf{h}$.
8. Use the line of best fit from step 6 to project the real-world simple model predictions $\underline{\mathbf{s}}_\mathbf{h}$ into real-world complex model predictions $\underline{\hat{\mathbf{s}}}_\mathbf{h}$ (thus correcting for potential bias).

Doherty and Christensen (2011) end their analysis here, as they assume a complex model so big that it can not feasibly be calibrated against the real-world data set. Furthermore, their correction of the real-world simple model prediction $\underline{\mathbf{s}}_\mathbf{h}$ via the regression line from step 6 equates the complex model to the (synthetic) real-world. This means that they treat the scatter discerned in step 6 as the (total) predictive uncertainty, and the real-world measurement data set $\mathbf{h}$ is used in steps 7 and 8 as a basis for bias correction.

In our application, while the complex model is deemed to be a reasonable representation of the real-world, it is imperfect, and thus prone to bias and/or error. Then again, it is still feasible to calibrate the model. Furthermore, we undertake the scatter plot analysis for some model outputs for which actual real-world measurements of the predicted model output are available. This real-world data corresponding to a prediction, h, can be used to estimate ϵ_{CM}, the prediction-specific error between CM and reality. Therefore, we altered and extended the above methodology to incorporate this new definition of the complex model as an error-prone simplification of reality, diverting after step 6 of the original method by Doherty and Christensen (2011). Note that a thorough analysis of these changes follows in Section 2.5.2.

Table 1
Notation of the different predictions made by the complex model (CM) and simple model (ROM) when generated by random CM realizations (s_i) or when calibrated on different data sets: (a) the real-world measurements $\mathbf{h}$, (b) the ith random CM output $\mathbf{o}(i)$ and (c) the real-world calibrated CM output $\mathbf{o}(\mathbf{h})$. The notation for $j = 1, \dots, m$ different predictions s^j is omitted here for clarity.

		Data set basis				
		$\mathbf{h}$	$\mathbf{o}(i)$	$\mathbf{o}(\mathbf{h})$		
Predictor	Observation	h	–	–		
	CM	$s_\mathbf{h}$	s_i	–		
	ROM	$\underline{s}_\mathbf{h}$	$\underline{s}_i$	$\underline{s}_{o(\mathbf{h})}$	bias correction →	$\underline{\hat{s}}_{o(\mathbf{h})}$

Table 1 summarizes the notation used for the different predictions made by the complex (CM) and simple (ROM) model.

7. (new) Calibrate the complex model $\mathbf{Z}$ against the real-world measurement data set $\mathbf{h}$. This creates a complex model output $\mathbf{o}(\mathbf{h})$ and calibrated complex model predictions $s_\mathbf{h}$.
8. (new) Calibrate the simple model $\mathbf{Z}_s$ against the calibrated complex model output $\mathbf{o}(\mathbf{h})$ and compute the corresponding model predictions $\underline{s}_{\mathbf{o}(\mathbf{h})}$. These predictions are the basis for the simplification error analysis (tested via the $[\underline{s}_{\mathbf{o}(\mathbf{h})}; s_\mathbf{h}]$ pairs) and bias correction, which will be explained in detail in Section 2.5.2.
9. (new) The line of best fit from step 6 is used to project the predictions $\underline{s}_{\mathbf{o}(\mathbf{h})}$ of the simple model (calibrated against the output $\mathbf{o}(\mathbf{h})$ of the real-world-calibrated complex model), into corresponding simple model predictions $\underline{\hat{s}}_{\mathbf{o}(\mathbf{h})}$ (corrected for bias by model simplification). This is shown in Fig. 2(b). Again, we differ from the original steps 7 and 8 by correcting $\underline{s}_{\mathbf{o}(\mathbf{h})}$ and not $\underline{s}_\mathbf{h}$.
10. (new) (originally step 7) Calibrate the simple model $\mathbf{Z}_s$ against the real-world measurement data set $\mathbf{h}$ and compute the corresponding model predictions $\underline{s}_\mathbf{h}$.
11. (new) For the j predictions where real-world data corresponding to predictions, h, is available, the data pairs of $[\underline{s}_\mathbf{h}^j; h^j]$ are added to the corresponding plots. $s_\mathbf{h}^j - h^j$ gives the discrepancy between the corresponding jth complex model prediction and reality, ϵ_{CM}. This is shown schematically in Fig. 2(c).

2.5.2. Analysis of the $\underline{s}$ versus s scatter plot

Section 2.4 described the theoretical basis for the scatter plots of paired simple complex models, while Section 2.5.1 outlined the application of this theory. As the resulting scatter plots are a major part of the analysis presented in this paper, we provide a detailed explanation of the different features and the information contained in such scatter plots in this section. The first part is based on Doherty and Christensen (2011) and is therefore only summarized here. The second part explains the changes and extensions of our approach that are listed in Section 2.5.1 in more detail.

Fig. 2(a) shows a schematic scatter plot of simple model predictions $\underline{s}$ versus complex model predictions s obtained by calibration of the simple model against n randomly parameterized complex model outputs that were corrupted by white noise to represent measurement error (labeled exemplarily for realization i). In addition, it depicts the regression

line of the scatter, it is 95% uncertainty bands and the reduction-based predictive error contributions (steps 1–6 in Section 2.5.1).

In general, the form and range of scatter points within the uncertainty bands and the slope and intercept of the regression line are very informative. The relationship between simple and complex model predictions is fitted by linear regression, from which we calculate the slope a and intercept b of the regression line. The intercept b denotes the bias of the ROM prediction (in comparison to the complex model prediction) that needs to be added (or subtracted) for correction of the prediction. The slope a can deviate from an ideal slope of 1 in two directions: if $a < 1$, the total horizontal extent of the scatter points is larger than the vertical scatter, which means that ROM predictions are more variable than (random) complex model predictions. This indicates that the propensity of predictive error was actually increased through calibration of the simple model compared to the random uncalibrated complex model (Doherty and Christensen, 2011). If, instead, $a > 1$, the total horizontal scatter is smaller than the total vertical scatter, and thus the random complex model predictions are more variable than the ROM predictions. This would mean that the ROM can not reproduce all possible predictive outcomes of the complex model, which might be an indication of over-fitting to the calibration data set or of missing process detail in the ROM.

As outlined in Section 2.4, the horizontal scatter around the line of regression is composed of the contributions of terms 2 and 3 of Eq. (21) to the simplification error of a simple model prediction $\underline{s}$, the influence of measurement noise (labeled ϵ_{Noise} in Fig. 2(a), which is equal to term 2) and the effect of parameter surrogacy (ϵ_{Surr} = term 3). The vertical scatter around the line of regression consists of terms 1 and 4 of Eq. (21), the null space contribution to the predictive error caused by the model reduction (ϵ_{Null}= term 1) and the influence of structural simplification on the ROM prediction (ϵ_{Redu} = term 4). All the above is explained in more detail in Doherty and Christensen (2011).

Now, we analyze the changes made to the original methodology and their impact on the analysis using scatter plots. Originally, Doherty and Christensen (2011) use $\underline{s}_{\mathbf{h}}$ for bias correction: the simple model prediction calibrated against the real-world data set **h**. As they state that the complex model is reality, they treat **h** as a complex model output and the regression line as a representation of the relationship between real world (represented by the complex model) and the simple model. Therefore, the regression line can be used to correct $\underline{s}_{\mathbf{h}}$ (old step 7).

We do not treat the complex model as reality (thus $\mathbf{o(h)} \neq \mathbf{h}$). The regression line depicts the bias between complex and simple model (since it was developed from calibrating the simple model against complex model outputs). Consequently, we use the regression line to correct $\underline{s}_{\mathbf{o(h)}}$, the prediction of the simple model calibrated against **o(h)** (instead of $\underline{s}_{\mathbf{h}}$, the prediction of the real-world calibrated simple model). **o(h)** is the output of the complex model calibrated against **h**, which contains any error and/or bias of the complex model. Therefore, $\underline{s}_{\mathbf{o(h)}}$ relates the simple and complex model, and can thus be used for correction of bias between simple and complex model (new steps 7 to 9) when using the simple model as a surrogate for the complex model. Fig. 2(b) shows the graphical analysis corresponding to the new steps 7–9 in the methodology in Section 2.5.1. The prediction $\underline{s}_{\mathbf{o(h)}}$ of the simple model calibrated against the (real-world calibrated) complex model output **o(h)** is added to the x-axis of the graph. From this prediction, two relationships can be drawn as shown in Fig. 2(b) and explained above. First, the regression line is used to correct the prediction bias denoted by the intercept b, which results in the corrected simple model prediction $\hat{\underline{s}}_{\mathbf{o(h)}}$ as shown in the y-axis. In Doherty and Christensen (2011), the vertical scatter around the intersection with the regression line (which is equal to $\epsilon_{\text{Redu}} + \epsilon_{\text{Null}}$, c.f. Fig. 2(b)) is equivalent to the postcalibration uncertainty of the corrected ROM prediction. In our case, the vertical scatter denotes the simplification error of the corrected ROM prediction $\hat{\underline{s}}_{\mathbf{o(h)}}$. Second, we calibrated the complex model (step 7 (new) in Section 2.4) against the real-world data set **h**. We add the corresponding (calibrated) complex model prediction $s_{\mathbf{h}}$ to the y-axis of the graph. The [$\underline{s}_{\mathbf{o(h)}}$; $s_{\mathbf{h}}$] pairs are now used to test the simplification error estimation for the corrected simple model prediction $\hat{\underline{s}}_{\mathbf{o(h)}}$. As we assume that the uncertainty bands depicted in the graph capture 95% of the simplification error of the ROM prediction $\hat{\underline{s}}_{\mathbf{o(h)}}$, the intercept between ROM prediction $\underline{s}_{\mathbf{o(h)}}$ and complex model prediction $s_{\mathbf{h}}$, i.e. the corresponding [$\underline{s}_{\mathbf{o(h)}}$; $s_{\mathbf{h}}$] pair, should ideally lie within these uncertainty bands. If the intersection lies far outside these uncertainty bands, that suggests the method of correcting the ROM prediction and assessing its error failed for the particular prediction. While non-uniqueness of calibrated CM parameters could influence [$\underline{s}_{\mathbf{o(h)}}$; $s_{\mathbf{h}}$] pairs, tests have shown only minimal effect for our chosen models (results not shown).

It is further possible to define "predictions" of data type and location for which one has actual measurements available. This could be the case for predictions of groundwater head during a validation phase, for example. If this is the case, these real-world measurements h^j taken at the same time and location as the corresponding model prediction s^j can be used to examine the performance of the complex model for this prediction. This is shown in Fig. 2(c) and corresponds to steps 10 and 11 (new) in the methodology in Section 2.4. While we did not use the prediction of the simple model calibrated against the real-world measurement set, $\underline{s}_{\mathbf{h}}$, for the bias correction as described above, it is now plotted on the x-axis of the graph against the real-world measurement h on the y-axis to create a new intersection, the [$\underline{s}_{\mathbf{h}}$; h] pair. If the complex model would perfectly represent reality, $h = s_{\mathbf{h}}$. Therefore, the difference on the y-axis, labeled ϵ_{CM} in Fig. 2(c), is an estimate of the prediction-specific error of the calibrated complex model against reality. Note that considering an (unknown) measurement error in h is generally possible and could potentially increase the error estimate between calibrated complex model and reality. Nonetheless, ϵ_{CM} can be considered here as a qualitative estimate of the predictive quality of the complex model.

To summarize the above, the graphical analysis of the paired simple complex models allows several insights:

- the slope a of the regression line is used to detect an increase of propensity of predictive error through calibration ($a < 1$) or potential over-fitting ($a > 1$),
- the intercept b of the regression line allows for bias-correction of the simple model prediction $\underline{s}_{\mathbf{o(h)}}$ to $\hat{\underline{s}}_{\mathbf{o(h)}}$,
- the uncertainty bands of the regression line are used to discern the influence of measurement noise and parameter surrogacy (horizontal scatter, terms 2 and 3 of Eq. (21)) as well as null space contribution and structural simplification (vertical scatter i.e. simplification error, terms 1 and 4 of Eq. (21)),
- the calibrated complex model prediction $s_{\mathbf{h}}$ allows the validation of the bias-correction via the [$\underline{s}_{\mathbf{o(h)}}$; $s_{\mathbf{h}}$] pairs and evaluation of predictive coverage with the uncertainty bands for the simple model prediction $\hat{\underline{s}}_{\mathbf{o(h)}}$ and
- potential real-world measurements h are used to discern the prediction-specific error ϵ_{CM} of the complex model prediction $s_{\mathbf{h}}$ ([$\underline{s}_{\mathbf{h}}$; h] pairs).

2.5.3. Comparison of full model and ROM parameters

The relationship between complex and simple model parameters was derived mathematically in Section 2.3. Here, we summarize the implications of these equations for the analysis of our results as previously discussed by Watson et al. (2013).

Eq. (14) can be used to calculate the composition of simple model parameters in regard to the complex model parameters. The main purpose of this analysis here, as was stated in Section 2.3, is the identification of parameter surrogacy (meaning that simple model parameters can take on the role of very different complex model parameters or processes). This is important given the fact that often, simple model parameters are thought to represent measurable real-world characteristics (e.g. hydraulic conductivity) or processes, which are incorporated into the model as initial values and parameter limits, for example. Thus, simple model parameters found to represent widely different complex

model parameters should be treated as "effective" parameters in model calibration and uncertainty analysis.

3. Case study: the Wairau aquifer model

3.1. The Wairau aquifer

In contrast to previous studies we applied the paired model analysis to a complex non-linear coupled surface water - groundwater interaction model. The study is conducted for the Wairau Plain aquifer, which is a shallow aquifer of regional importance on the Wairau Plain in the northern part of the South Island of New Zealand. Its main source of recharge is the Wairau River, and its groundwater is used intensively for irrigation (mainly vineyards) and municipal uses (drinking water for the city of Blenheim and surrounding villages).

The fluvial, highly conductive aquifer has formed under the approximately 27 km long floodplain of the Wairau River, a braided gravel bed river that discharges into the Pacific Ocean. The aquifer is shallow, has a thickness of 20–35 m and is horizontally stratified by three main hydrogeological units that comprise a lower and upper highly permeable gravel aquifer, intermitted by a lower-permeable clay-rich gravel layer. In the East, toward the sea, a fourth, confining geological formation, Dillon's Point, lies like a wedge on top of the first unit, forcing the groundwater upwards into several low-land springs and into the downstream part of the Wairau River. Several groundwater observation wells to measure groundwater head are installed on the Wairau Plain, and a flow-recorder monitors the discharge of the major low-land spring (Spring Creek). Furthermore, four river gauging stations are located in the Wairau River. Meteorological data (precipitation and potential evapotranspiration) were taken from NIWA's virtual climate station network (Tait et al., 2006). Estimates of irrigation water abstraction and groundwater recharge were calculated by a soil–water balance model (Wöhling et al., 2018). A detailed numerical model was developed for the coupled surface water–groundwater system which was recently presented in Wöhling et al. (2018). This model is used as the complex benchmark model in our paired model analysis and briefly summarized in the next section.

3.2. Complex MODFLOW (CM) model

The transient, numerical, coupled surface water - groundwater model was presented in detail by Wöhling et al. (2018) and is therefore only summarized here. It was implemented in MODFLOW-NWT (Niswonger et al., 2011) using a horizontal 200 × 200 m grid (approx. 85 km^2 and 6360 active model cells) and three vertical layers analogous to the geological units (Fig. 3(a)). The model was calibrated on about three years of data (1225 daily time steps). The northern model boundary follows the course of the Wairau River, while the southern model boundary was aligned orthogonal to the groundwater flow field and thus assigned a no-flow boundary. In the east, a constant groundwater flux boundary was implemented. The Wairau River is simulated using MODFLOW's stream-flow routing package (Niswonger and Prudic, 2005) allowing river and groundwater to interact in both directions over a head-gradient flow that is constrained by the river bed conductivity K_R. The low-land springs (Fig. 3(a)) are implemented as drains, allowing groundwater to leave the model along a conductivity (K_D) and gradient-controlled one-way interaction. At the land surface, the model area is subject to groundwater recharge estimated by a soil water balance model (Wöhling et al., 2018).

The model has a total of 207 parameters. The majority of parameters (180) is used for interpolating hydraulic conductivity K_H and specific yield S_y of the three geological layers via pilot points. The specific storage S_S and anisotropy ratio f_a are defined by a single parameter per geological unit, as are all four geological properties of the eastern Dillon's Point formation. Other parameters include river bed conductivities K_R as well as bed conductivities for the different low-land springs (K_D). A list of all CM model parameters and their respective ranges for calibration is given in Table 2. Spatially variable K_H and S_y parameter fields were implemented using regularization techniques and an exponential variogram.

The parameters were calibrated using groundwater heads (labeled 1–5), discharge at the Spring Creek (SC) spring, differential Wairau River flow gaugings and "soft" calibration targets for other springs that were implemented as expert knowledge. State of the art parameter estimation was performed using PEST (see Doherty, 2016) in Pareto mode, calibrating the model parameters and assessing the trade-off between data and regularization objective functions. The model is used to predict groundwater heads at locations not contained in the calibration data set (labeled 6–10), river–groundwater exchange fluxes and flows of other low-land springs. An in-depth discussion of the calibration process and uncertainty analysis for this model can be found in Wöhling et al. (2018) and is therefore not repeated here.

The run time of the complex MODFLOW (CM) Wairau model is between 75 and 150 s on an average desktop computer (i5 processor with 2.6 GHz, 16 GB RAM, and a 64 bit operating system). This seems low, but considering the high amount of runs necessary for calibration (208 for each computation of the Jacobian matrix), the computational burden for stochastic simulation is rather high. It also allows for the calibration of the complex model which is a key feature of the extended paired model analysis presented here. We like to note that model complexity is not necessarily directly synonymous with the number of model cells or parameters (cf. Höge et al., 2018; Schöniger et al., 2014). It has been shown that linear models with 10^4 model cells can efficiently be replaced by ROM's (e.g. Siade et al., 2012). Non-linear models, such as the surface water–groundwater model presented here, and state-depending boundaries, are much more challenging to reproduce for ROMs.

For our analysis, a total of 1000 random parameter sets were generated and model simulation conducted, of which 976 terminated successfully and were used for the paired model analysis. For all non-pilot point parameters, the values were drawn from a uniform distribution of their respective parameter ranges (Table 2). For the pilot point parameters of the three aquifer layers (K_H and S_y fields) we developed a method to semi-randomly generate parameter fields which adhere to both their respective parameter ranges and the spatial correlation of the parameters. A detailed description of this method can be found in Appendix A.

3.3. Simplified MODFLOW (SM) model

The first ROM used in this study is a strongly simplified MODFLOW model that is subsequently referred to as the SM model. It is based on the CM model described in Section 3.2 and uses identical boundary conditions. The horizontal grid size was changed from 200 × 200 m to 600 × 600 m, thus reducing the number of active cells from 6360 to 756. Furthermore, the K_H and S_y fields of the CM model were assumed to be uniform in each hydrogeological unit and thus replaced by a single parameter each. In addition, the number of river bed sections was reduced from 12 to 3. As a consequence, the total number of SM model parameters is 24, compared to the 207 parameters of the CM model. Fig. 3(b) depicts the differences of the computational grid and Table 2 lists how the CM and SM model parameters are related.

The calibration process for the SM model, along with the utilized calibration data and weighting strategy, were kept consistent with the strategy used for the CM model (see Wöhling et al., 2018, for details). The model simplification resulted in a reduction of run time for a single SM model run by a factor of 5–10 (approx. 15 s for a single SM model run). In addition, the drastic parameter reduction makes calibration or stochastic simulation with the SM model much less computational demanding than with the CM model (25 instead of 208 model runs for computing the Jacobian matrix).

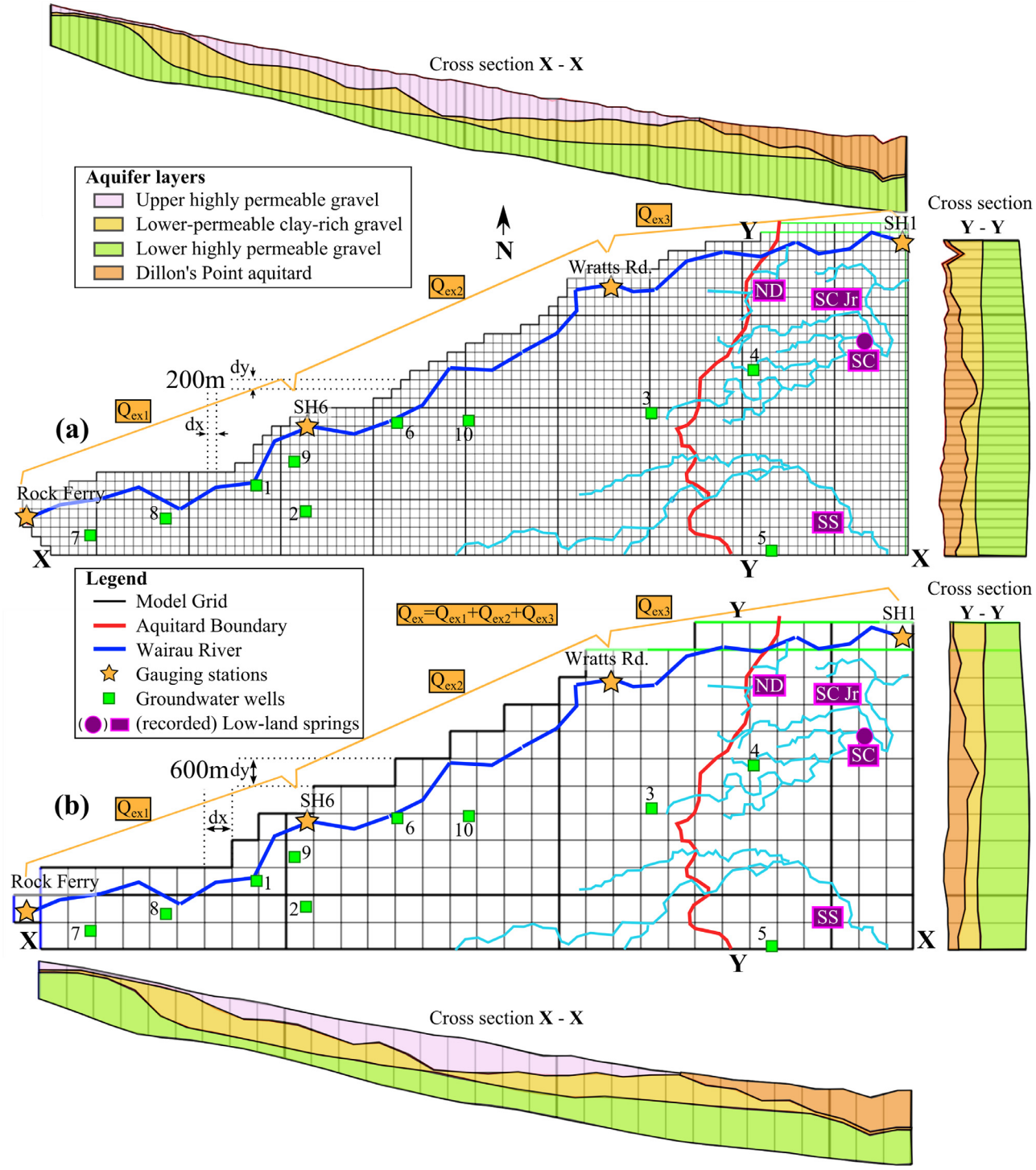

Fig. 3. (a) Complex CM and (b) simple SM MODFLOW models with the main geological features.

3.4. Artificial neural network (ANN)

The second type of ROM used in this study are artificial neural networks (ANNs). We created separate ANNs for each of the data time series described in Section 3.5 (with exception of the total river–groundwater exchange flux, Q_{ex}) instead of a single ANN model simulating all the different data series as the CM and SM models do. Thus, we used 17 ANNs, one for each data time series corresponding to a prediction used in the scatter plot analysis. Their basic configuration is a nonlinear regressive neural network that was built using the Neural Network Toolbox of MATLAB© (Version R2015a). We used the river flow time series estimated for the Rock Ferry gauging station as well as the groundwater abstraction and recharge time series integrated over the entire model area as input to all ANNs. Furthermore we implemented a closed feedback loop where the simulated outputs from previous time steps are used as input. The time steps used for the ANNs are the same as for the CM and SM models. Each individual ANN was trained on the continuous

Table 2
Parameterization of the Wairau MODFLOW models.

Parameter			CM model		SM model	
Type [unit]	Property	Range	Type	Quantity	Type	Quantity
Hydraulic conductivity (K_H) [m/d]	Layer 1	1–3500	Pilot points	26	Zones	1
	Layer 2	1–2500	Pilot points	31	Zones	1
	Layer 3	1–3500	Pilot points	33	Zones	1
	Aquitard	0.1–50	Zones	1	Zones	1
Specific yield (S_y) [–]	Layer 1	1×10^{-3}–0.35	Pilot points	26	Zones	1
	Layer 2	1×10^{-3}–0.35	Pilot points	31	Zones	1
	Layer 3	1×10^{-3}–0.35	Pilot points	33	Zones	1
	Aquitard	1×10^{-7}–1×10^{-3}	Zones	1	Zones	1
Specific storage (S_S) [m^{-1}]	Each	1×10^{-7}–1×10^{-3}	Zones	4	Zones	4
Anisotropy ratio (f_a) [–]	Each	1–10	Zones	4	Zones	4
River bed conductivity (K_R) [m/d]	–	$1 \times 10^{-3} - 0.2$	Sections	12	Sections	3
Drain bed conductivity (K_D) [m/d]	–	$1 \times 10^{-4} - 1,000$	Sections	5	Sections	5
Total				207		24

Table 3
Structural configuration of the ANN models.

Datatype	ANN data set	Number of nodes	Input delay	Feedback delay
Groundwater heads	1: Conders Recharge	5	[1, 2, 3, 4]	[1, 2, 3, 4, 5, 6, 7, 8]
	2: Conders No. 2	20	[1, 2, 3]	[1, 2, 3, 4, 5, 6, 7, 8]
	3: Wratts Rd.	10	[0]	[1, 2, 3, 4, 5, 6, 7, 8]
	4: Selmes Rd.	10	[0, 1]	[1, 2, 3, 4, 5, 6, 7, 8]
	5: Murphys Rd.	5	[0]	[1, 2, 3, 4, 5, 6]
	6: Pauls Rd.	5	[0]	[1, 2, 3, 4, 5, 6]
	7: Old MCB	10	[1, 2]	[1, 2, 3, 4, 5, 6]
	8: MCB	10	[0]	[1, 2, 3, 4, 5, 6, 7, 8]
	9: Catchment Shallow	5	[1, 2, 3, 4]	[1, 2, 3, 4, 5, 6, 7]
	10: P Neal	5	[1, 2]	[1, 2, 3, 4, 5, 6, 7, 8]
River exchange fluxes	1: Q_{ex1}	5	[0, 1]	[1]
	2: Q_{ex2}	5	[0, 1]	[1, 2, 3, 4, 5, 6, 7, 8]
	3: Q_{ex3}	5	[0, 1]	[1, 2, 3, 4, 5]
Low-land springs	1: Q_{sp1}	5	[0]	[1, 2, 3, 4, 5, 6, 7, 8]
	2: Q_{sp2}	5	[0]	[1, 2, 3, 4, 5, 6, 7, 8]
	3: Q_{sp3}	10	[0, 1]	[1, 2, 3, 4, 5, 6, 7]
	4: Q_{sp4}	5	[0]	[1, 2, 3, 4, 5, 6, 7]

time series over the calibration time (days 124–1047) of its corresponding data type and location.

The three main structural components of the ANNs used in this study are the number of nodes in the hidden layer, the amount of input delay and the amount of feedback delay, all of which were chosen for each net individually based on a preliminary data-specific performance test with different ANN setups (results not shown). The feedback delay denotes the number of time steps of previous simulated outputs that are used in the closed feedback loop, while the input delay describes the time lag used for the input series. Note that the setup of separate ANNs for each of the data time series is possible in this paired model analysis, as full time series are available for each data point (since they are simulations of the CM model), while there are no corresponding real-world time series for a lot of these data points (see Section 3.5 for details). Table 3 summarizes the ANN structures that were used for each data set.

A single ANN model run takes about 0.1 s on the desktop PC.

3.5. Predictions

The paired model analysis presented in Section 2.5 is applied to several model predictions which are subsequently described.

The first group of predictions is a hindcast of groundwater heads at ten observation wells (Fig. 3). For each groundwater well, the smallest daily value in the whole simulation period was chosen as the prediction (labeled $h_{\min}$). Available measurements of the first five groundwater wells (1–5: Conders Recharge, Conders No. 2, Wratts Rd., Selmes Rd. and Murphys Rd.) were used in the calibration data set, therefore these are not model predictions in the true sense. The other groundwater wells (6–10: Pauls Rd., Old MCB, MCB, Catchment Shallow and P Neal) are model predictions of groundwater heads for which no data was used in the calibration process.

The second group of predictions are the river–groundwater exchange fluxes between two consecutive gauging stations (Q_{ex1}: Rock Ferry—SH6, Q_{ex2}: SH6—Wratts Rd. and Q_{ex3}: Wratts Rd.—SH1). A specific low-flow period was chosen in which the average exchange flux between river and groundwater in the respective river sections was measured. Furthermore, the overall mean exchange flux (Q_{ex}: Rock Ferry—SH1) in the model area (consisting of the sum of all three individual fluxes) during this time period was also used as a prediction for the SM model.

The third prediction group are the mean flows of the low-land springs. Flow data are available for a portion of Spring Creek which was used in the calibration data set of the model (labeled SC in Fig. 3 with predicted mean flow Q_{sp1}). The fluxes in the remaining tributaries of Spring Creek were summarized and used as a prediction under the name Spring Creek Jr. (SC Jr, Q_{sp2}) in this study. Furthermore, the various small southern streams (SS, Q_{sp3}) were summed up and were used as another flow prediction. Finally, the model predicts the total flow in a smaller spring that flows back into the river in the north-east of the model domain named Northern Drain (ND, Q_{sp4}).

4. Results

4.1. Paired model analysis: predictions

In this chapter we present the results of predictive performance and simplification error analysis for the two ROM types (Section 3.5) as analyzed using the paired model analysis (Section 2.5.2). We will com-

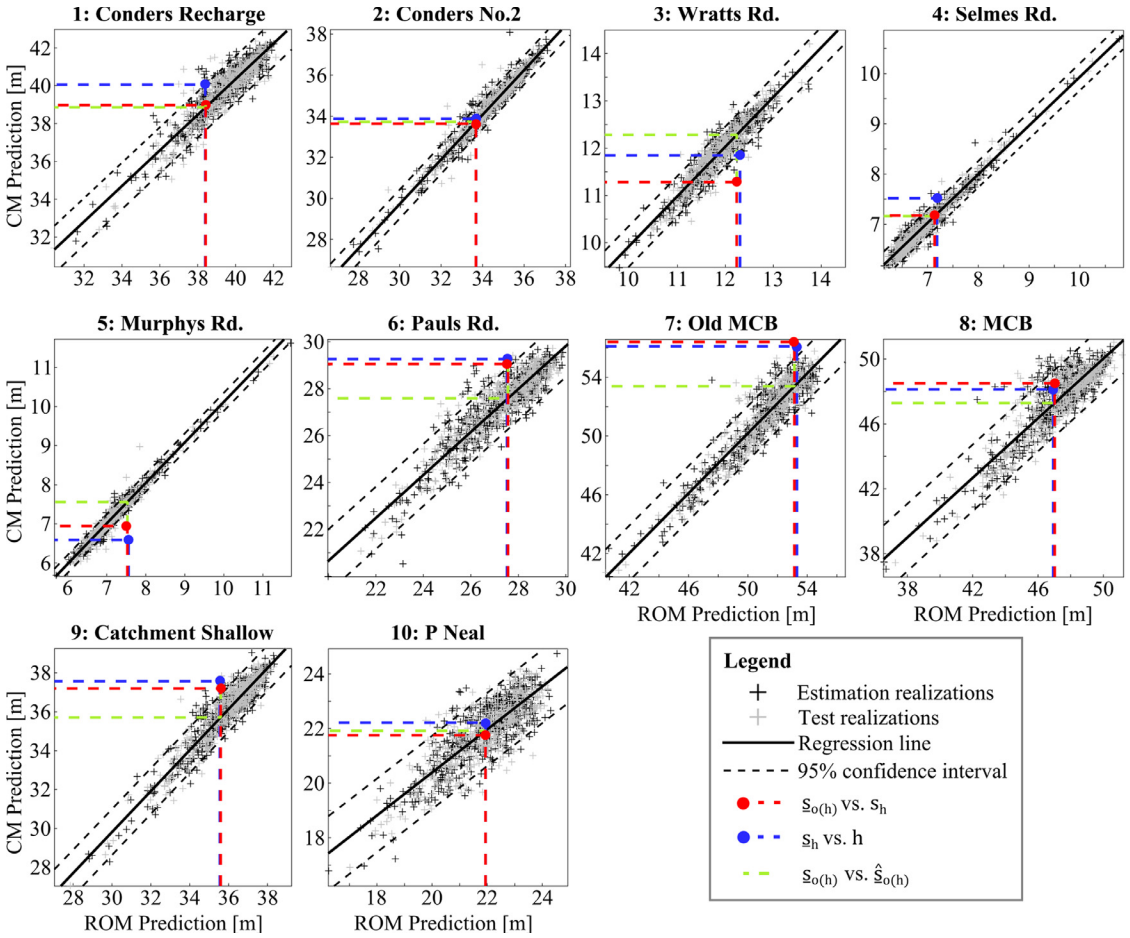

Fig. 4. Scatter plots of minimum groundwater head predictions of the SM model. Blue headers denote predictions pertaining to wells used in the calibration data set. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 4
Minimum groundwater head predictions h_{min} (Figs. 4 and 5): coefficients of regression (slope a and intercept b) and its statistics (coefficient of determination R^2 and standard deviation σ).

Prediction	SM				ANN			
	a	b	R^2	σ	a	b	R^2	σ
1: Conders Recharge	0.94	2.72	0.861	0.606	1.00	−0.32	0.985	0.200
2: Conders No. 2	1.09	−2.94	0.953	0.348	0.99	0.37	0.976	0.247
3: Wratts Rd.	1.05	−0.53	0.887	0.212	1.01	−0.29	0.981	0.086
4: Selmes Rd.	0.97	0.24	0.940	0.130	0.97	0.14	0.996	0.034
5: Murphys Rd.	1.04	−0.24	0.973	0.114	1.00	−0.03	0.995	0.049
6: Pauls Rd.	0.92	2.33	0.835	0.658	1.00	−0.09	0.977	0.244
7: Old MCB	1.02	−0.89	0.878	0.927	1.00	−0.14	0.990	0.262
8: MCB	0.92	4.04	0.795	1.020	1.04	−1.93	0.987	0.255
9: Catchment Shallow	1.05	−1.82	0.889	0.583	1.01	−0.42	0.983	0.230
10: P Neal	0.79	4.66	0.712	0.668	1.03	−0.70	0.972	0.208

pare the predictive quality of the two different ROM types for different predictions to evaluate their strengths and weaknesses.

The 976 total data points (paired complex-simple model runs) were divided into two subgroups. The first 500, named estimation realizations, were used to fit a regression line with 95% confidence intervals, while the remaining 476, named test realizations, were used to test the integrity of these confidence intervals. This allows to assess the quality of the regression as well as the structure of the residuals.

4.1.1. Groundwater head predictions

Fig. 4 shows the scatter plots of the SM model for the minimum groundwater head predictions h_{min} of the ten wells in our analysis. Intercept, slope and basic statistics of the regression lines are listed in Table 4. First of all, the h_{min} used in the calibration (1–5) exhibit narrower bands of scatter than those solely used for prediction (6–10), which is confirmed by on average higher R^2 and lower σ values (Table 4). This seems plausible, as predictions of data contained in the calibration data set should have smaller errors associated with them, as is predicated by the narrower scatter bands. The slope of the regression line is close to 1 for all h_{min} predictions, even where the intercept is far from 0. This means that while the SM model might be strongly biased (as indicated by the intercepts), one could assume that the correction of the prediction via the regression line is trustworthy. This is evaluated utilizing the [$s_{o(h)}$; s_h] pairs, which should lie within or close to the uncertainty bands.

Fig. 4 shows that this is not always the case. For example, the Wratts Rd. and Murphys Rd. h_{min} predictions (Fig. 4-3 and 5), the [$\underline{s}_{o(h)}$; s_h] pairs lie far below the uncertainty bands, indicating that the error of the model reduction is underestimated by the method. This discrepancy can be explained by the following: the Murphys Rd. well is influenced by the no-flow boundary because it is located on the southern boundary of the model area (well 5 in Fig. 3). Boundary effects are probably intensified by the grid coarsening, affecting a larger area in the SM model compared to the CM model. This is confirmed by the [$\underline{s}_h$; h] pair, which shows an even smaller real-world measurement h for the Murphys Rd. h_{min} than the CM model prediction s_h, indicating that (a) the CM model also shows an influence of the no-flow boundary on this prediction and (b) the amplification of this effect through the simplification. The Wratts Rd. well lies at the outcropping of the Dillon's Point aquitard in the eastern part of the model, where most of the low-land springs originate. This is a model area that has shown to be sensitive to changes in the model geometry, relying on a relatively detailed resolution of geology. This detail has probably been lost in the SM model. The [$\underline{s}_h$; h] pair suggests that the CM model itself is erroneous for the Wratts Rd. h_{min} prediction, which might exaggerate the amount of simplification error underestimation compared to the [$\underline{s}_{o(h)}$; s_h] pair.

In contrast to the Wratts Rd. and Murphys Rd. predictions, the [$\underline{s}_{o(h)}$; s_h] pair for the Old MCB well lies both above the uncertainty bounds and at a very high value of the random realizations (Fig. 4-7). Most of the test realizations are also found in this area with a bias toward CM model values significantly higher than their corresponding SM model values for these high Old MCB h_{min} prediction values. This indicates that for these values, a linear regression line might not be a good estimator for model reduction error, and an upward-curving non-linear fit might be more suitable to capture the simplification error for high predictive values. This is also the case at the Pauls Rd. and MCB wells (Fig. 4-6 and 8).

Fig. 5 shows the h_{min} prediction scatter plots for the ANNs, and the statistics of the linear regressions are given in Table 4. Generally, the slope of the best fit line is very close to 1 for all predictions, and the uncertainty bounds are quite narrow. The [$\underline{s}_{o(h)}$; s_h] pairs are within the uncertainty bands for almost all head predictions, and reasonably close for the remaining ones. As each individual ANN is calibrated against the corresponding head time series of the complex model (see Section 3.4), [$\underline{s}_h$; h] pairs could only be computed for the wells where a real-world data set was available (wells 1–5). For these cases, the [$\underline{s}_h$; h] pairs show little influence of the prediction-specific error between CM model and real-world on the ANN model predictions.

For the Old MCB and MCB wells (Fig. 5-7 and 8), the clustering of test realization scatter below the uncertainty band for low prediction values suggests that a linear relationship might not be suited to transfer the ANN prediction to the CM predictions. As such low h_{min} predictions are rare cases at these wells and the real-world calibrated CM and ANN prediction pairs [$\underline{s}_{o(h)}$; $\hat{s}$] are at higher values, this seems of not much significance for the presented application.

In comparison to the SM model predictions of h_{min}, the ANN model predictions are less biased and show far less simplification error. This is mainly due to the different setup, though, as explained in Sections 3.3–3.5. The SM model is calibrated against a complex data set comprising several groundwater head time series (1–5) as well as spring flows and river exchange fluxes and is used to predict simultaneous system states contained and not contained in the calibration data set. In contrast, the ANNs are specific to their respective system state and solely calibrated to the single time series of the specific predictions. Due to this specialization, the ANNs are better equipped to predict h_{min} than the SM model, but they can only be applied where real-world data sets are available.

4.1.2. River–groundwater exchange flux predictions

The scatter plots for the prediction of river–groundwater exchange fluxes for the SM model are presented in Fig. 6, while the statistics of the linear regressions are listed in Table 5. Again, the most apparent feature is the difference between the two predictions pertaining to the calibration data set (Q_{ex3} and Q_{ex}) and the other two predictions (Q_{ex1} and Q_{ex2}).

Table 5
River–groundwater exchange flux predictions Q_{ex} (Figs. 6 and 7): coefficients of regression (slope a and intercept b) and its statistics (coefficient of determination R^2 and standard deviation σ).

Prediction	SM				ANN			
	a	b	R^2	σ	a	b	R^2	σ
1: Q_{ex1}	0.40	1.69	0.315	0.459	0.93	0.06	0.990	0.056
2: Q_{ex2}	0.45	1.67	0.320	0.515	0.94	0.01	0.991	0.060
3: Q_{ex3}	1.01	0.06	0.911	0.110	0.97	−0.05	0.994	0.029
4: Q_{ex}	0.94	0.42	0.939	0.200	–	–	–	–

Table 6
Low-land spring flow predictions Q_{sp} (Figs. 8 and 9): coefficients of regression (slope a and intercept b) and its statistics (coefficient of determination R^2 and standard deviation σ).

Prediction	SM				ANN			
	a	b	R^2	σ	a	b	R^2	σ
1: Q_{sp1}	1.01	0.01	0.980	0.166	0.99	0.03	0.997	0.048
2: Q_{sp2}	0.76	−0.27	0.770	0.273	0.99	0.00	1.000	0.010
3: Q_{sp3}	0.64	−1.00	0.596	0.322	0.99	0.02	0.991	0.049
4: Q_{sp4}	0.21	−0.18	0.093	0.233	0.98	0.01	0.997	0.012

As seen in Fig. 6-3 and 4, the scatter for the predictions Q_{ex3} and Q_{ex} is minorly spread, the slopes of the regression lines are close to 1 and the [$\underline{s}_{o(h)}$; s_h] pairs are inside the uncertainty bounds. The two available [$\underline{s}_h$; h] pairs show that there is little error between CM model and reality for those two specific predictions. For Q_{ex}, the scatter points outside the uncertainty bands are almost exclusively situated above the uncertainty band (Fig. 6-4), meaning there is a bias of the SM model underestimating these fluxes compared to the CM model.

The scatter plots of Q_{ex1} and Q_{ex2} look very different (Fig. 6-1 and 2), with wide-spread scatter and slopes below 0.5. As was explained in Section 2.5.2, this indicates an inability of the calibrated SM model to reduce the simplification predictive error in comparison to the precalibration uncertainty of the CM model. The large horizontal scatter implies that this is influenced by parameter surrogacy in the SM model. The river flow gauging stations corresponding to the calculations of Q_{ex1} and Q_{ex2} are situated in the north-western and northern area of the model (Fig. 3), where the river is in contact with the main aquifer. The hydrogeological aquifer properties are parameterized via interpolation between pilot points in the CM model, which is reduced to a single parameter in the SM model. This reinforces the necessity of representing hydrogeological detail which is only in part compensated by other parameters in the SM model. This is examined in more detail in Section 4.2. The [$\underline{s}_{o(h)}$; s_h] pairs in Fig. 6-1 and 2 suggest a strong underestimation of the simplification error for these two predictions, which further illustrates the poor ability to transfer the ROM predictions to the CM predictions.

Only three ANNs were built to predict the consecutive partial exchange fluxes between the gauging stations, and the results are shown in Fig. 7. For these three, slopes of the regression line are very close to 1 and uncertainty bounds are small while also incorporating [$\underline{s}_{o(h)}$; s_h] pairs. Furthermore, intercepts are close to 0 for all three predictions, meaning the ANNs can reproduce the CM model predictions without any notable bias (Table 5). As noted above, this is again due to the prediction-specific setup of the ANNs as ROMs.

4.1.3. Low-land spring predictions

Fig. 8 shows the scatter plots for the SM model predictions of mean low-land spring fluxes and the statistics for the corresponding regression lines are given in Table 6. Observation of spring flow is only available for Spring Creek.

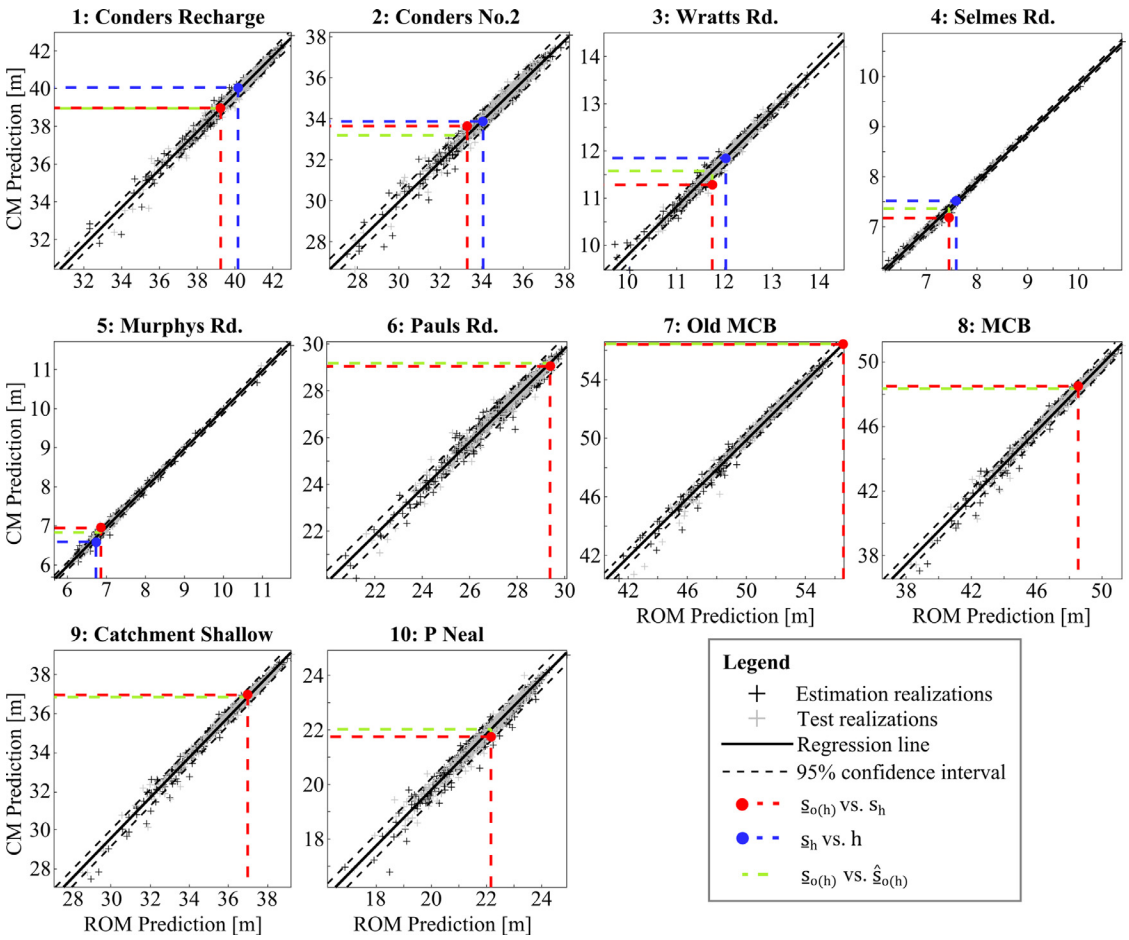

Fig. 5. Scatter plots of minimum groundwater head predictions of the ANN models.

The SM model shows almost zero bias for the prediction of Q_{sp1} (Fig. 8-1), and the regression line with slope about 1 and small corresponding uncertainty bands suggest only a small model simplification error for this prediction.

For both Q_{sp2} and Q_{sp3}, the regression lines have a slope of about 0.7 and wider uncertainty bands (Fig. 8-2 and 3). Nonetheless, the [$\underline{s}_{o(h)}$; s_h] pairs are captured by the simplification error bounds for both predictions. The [$\underline{s}_h$; h] pair is far outside the uncertainty bands for Q_{sp3} (Fig. 8-3), showing a huge discrepancy between the observation and the calibrated CM model prediction. This suggests that while the regression fit seems reasonable, the calibrated CM model prediction seems to be a special case that might restrict the model pairing for this particular prediction. On the other hand, the observation o_s for Q_{sp3} itself is known to be a very uncertain estimate, so it might not be the best indicator for predictive quality.

The SM model prediction of Q_{sp4} shows a scatter plot distinctly different to all others (Fig. 8-4). A multitude of realizations predict a Q_{sp4} of zero, some in agreement between CM and SM model predictions, but many only in the SM model. Many model realizations seem to predict groundwater levels below the threshold necessary to supply the spring with flow (i.e. below the spring's channel elevation). These realizations predict flows of $Q_{sp4} = 0$ m^3/s which results in this particular scatter plot that allows no meaningful analysis of the results. This is demonstrated by the [$\underline{s}_h$; h] pair, which is one of the few points where the SM model predicts $Q_{sp4} > 0$ m^3/s while the accompanying CM model predicts $Q_{sp4} = 0$ m^3/s (Fig. 8-4). The accumulation of SM model realizations of $Q_{sp4} = 0$ m^3/s compared to the few CM model realizations suggests a significant error in model simplification. The inability of the SM model to reproduce small-scale interactions between groundwater and surface features in this model area results in either (a) too low groundwater levels in the SM model or (b) overestimation of spring bed elevation through spatial aggregation.

Similar to the results presented for h_{min} and Q_{ex} in the previous section, the results for spring predictions with ANNs are different from those of the SM model. The scatter and uncertainty bounds for the ANN models around the regression lines shown in Fig. 9 can barely be distinguished, with slopes close to 1 and intercepts around zero (Table 6). The [$\underline{s}_{o(h)}$; s_h] pairs lie within the uncertainty bands for all four predictions. Note again that a direct comparison to the SM model results is not informative, as the ANN models were each calibrated against it is individual specific data set. Nonetheless, the ANN models seem to be able to predict mean spring fluxes for all four springs with small error and without any need for correction through a complex model.

4.2. Composition of SM model parameters

The equations presented in Section 2.3 allow an analysis of the effects of parameter reduction on the SM model parameters regarding their composition of CM model parameters. Note that the parameters of the ANNs are not associated to any real-world features or characteristics of the aquifer (in contrast to the SM model parameters) and we therefore only carried out the analysis for the SM model, even though it would be possible for the ANNs as well.

As can be seen in Table 2, some of the 24 SM model parameters are combinations of the 207 CM model parameters (the hydraulic conductivity and specific yield of layer 1 to 3 as well as the river bed conductivities), while others are the same (the remaining ones). All of the CM model parameters stand for certain processes of the real-world geology

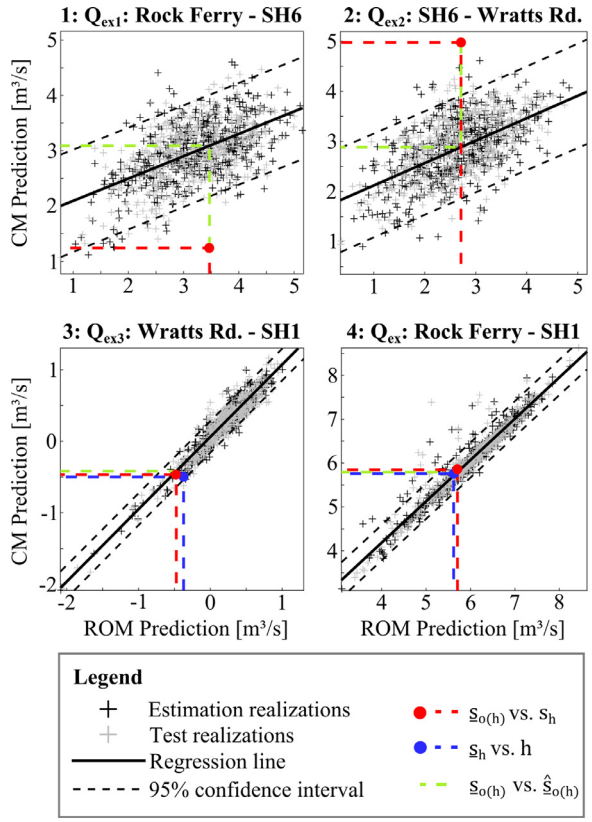

Fig. 6. Scatter plots of low-flow period river–groundwater exchange flux predictions of the SM model. Blue headers denote predictions pertaining to exchange fluxes used in the calibration data set. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

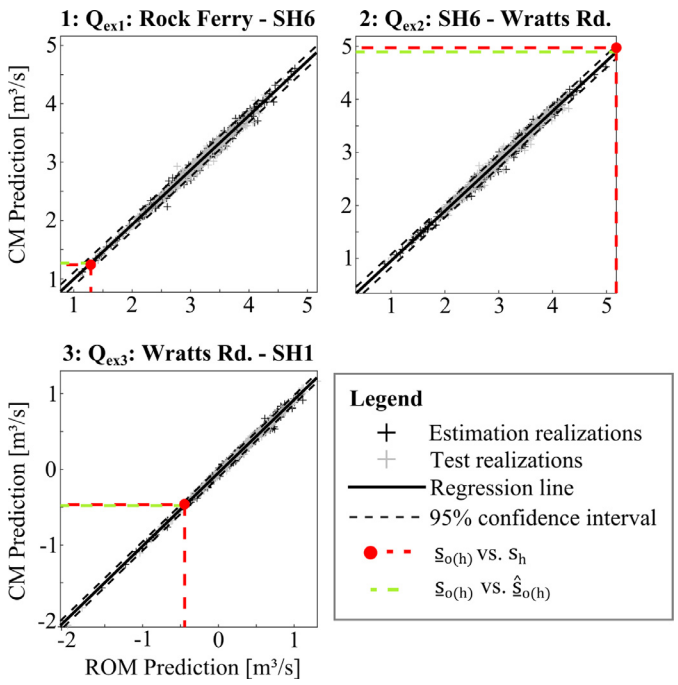

Fig. 7. Scatter plots of low-flow period river–groundwater exchange flux predictions of the ANN models.

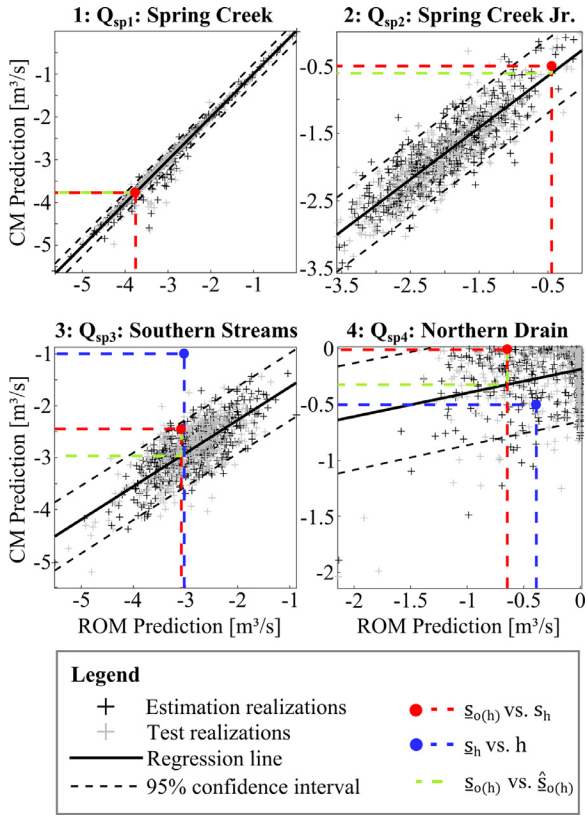

Fig. 8. Scatter plots of mean spring flow predictions of the SM model. Blue headers denote predictions pertaining to spring fluxes used in the calibration data set. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

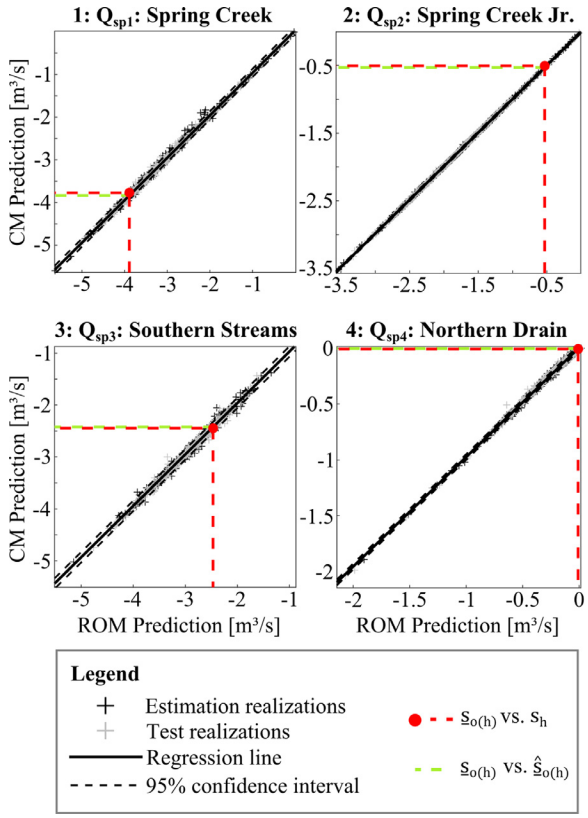

Fig. 9. Scatter plots of mean spring flow predictions of the ANN models.

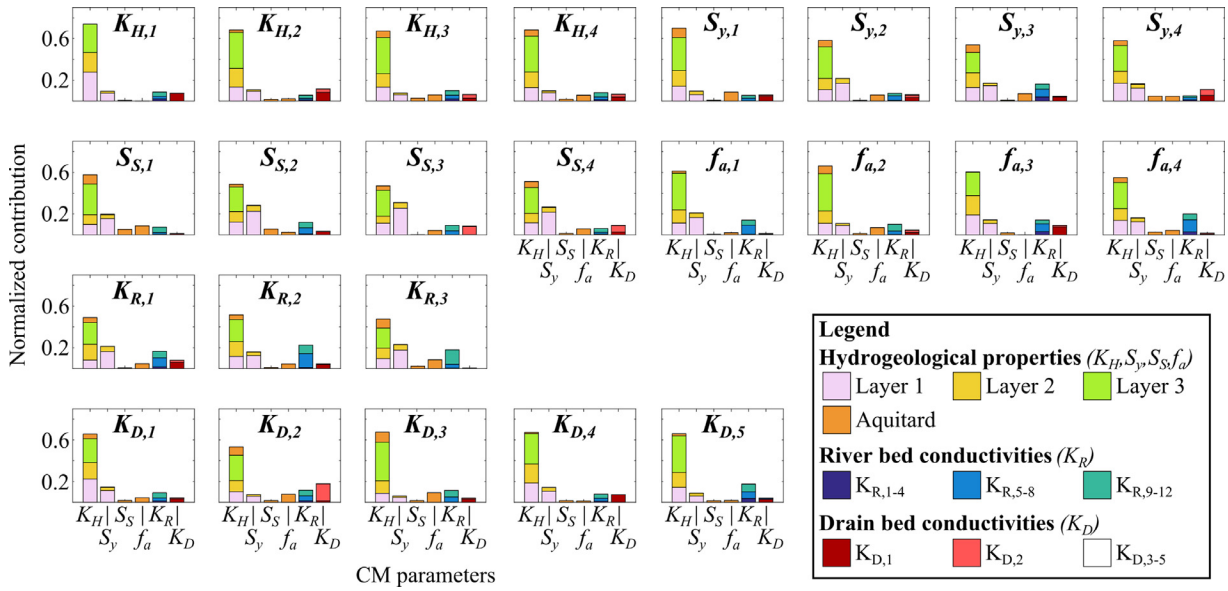

Fig. 10. Normalized contribution of the (grouped) complex model parameters to the 24 simple SM model parameters.

and were thus informed by expert knowledge to derive starting values and parameter bounds for the calibration. Fig. 10 shows the contribution of the CM model parameters to each of the (grouped) SM model parameters. Preferably, a SM model parameter should be mainly influenced by the corresponding CM model parameters ($K^{SM}_{H,1}$ by K^{CM}_{H} layer 1 pilot points, for example). It is generally notable that many different CM model parameters contribute to each SM model parameter. Only in some cases are the CM model parameters corresponding to the same process a major influence on the accompanying SM model parameter, as for $K^{SM}_{D,2}$ and, to some extent, $S^{SM}_{S,2}$ or $K^{SM}_{H,1}$. From the discussion in Section 4.1.2, we assume a high level of parameter surrogacy for $K^{SM}_{D,1-2}$, with major influences of K^{CM}_{H}, which can clearly be seen in the graph. Furthermore, $K^{SM}_{R,1}$, between gauging stations Rock Ferry and SH6, does not have any real contributions from its CM model counterparts, $K^{CM}_{R,1-4}$.

This is a crucial demonstration of the concept of parameter surrogacy that was mentioned in several sections of this study. Fig. 10 demonstrates clearly that parameters in the physically-based, distributed SM model are not only, or, in many cases even barely, representing the function of the real-world process associated with them. Instead, they are made up of various mixtures of complex CM model parameters. This does illustrate that the information of starting values and parameter bounds via expert knowledge is dubious at best for this ROM. Indeed, the limiting of parameter values to geology-informed delimitation might severely inhibit the ROM performance in a case where parameters are not representing the process that one would associate with them. Therefore, ROM parameters should best be treated as “effective” parameters and only be loosely constrained in application.

5. Discussion and conclusions

In this study, the theoretical work of Doherty and Christensen (2011) and Watson et al. (2013) of a paired model analysis of simplification error for ROMs was extended and applied to the real-world, physically-based groundwater model of the Wairau Plains Aquifer. Two ROMs differing widely in type and complexity were created for the paired model analysis, the first a simplification of the complex MODFLOW model, the second a collection of ANNs. The trustworthiness of the predictive bias correction and simplification error assessment that is the focal point of the scatter plots was tested by actual calibration of the complex model. In addition, the application to two types of ROMs demonstrated potencies and shortcomings of these ROMs for different application fields.

The first ROM, labeled SM, is a spatially and parametrically simplified version of the basic complex model, CM. The proposed method of bias correction and error analysis seems to work fine for most of the predictions tested with this ROM, with low errors associated with predictions pertaining to the calibration data set, medium errors for predictions of similar type, and high errors for the remaining predictions. Nonetheless, the testing of the analysis via the real-world calibrated complex CM model prediction showed that the error estimates for some predictions, disregarding of type, are too low. These problems seem to have several reasons. A major one seems to be the location of the prediction. Several “problematic” predictions lie within the area of transition between the highly conductive aquifer and the Dillon’s Point aquitard in the East of the model. This area requires relatively high spatial and parametric detail to be adequately represented which seems to be missing in the SM model. Another source of simplification error underestimation seems to be coming from parameter surrogacy, which could be the main contributor for the two highly uncertain river–groundwater flux predictions. As demonstrated by our analysis, parameters in the SM model compensate for an assortment of CM model parameters due to absent variability and process representation in the simple model. This questions the constraining of simple model parameter boundaries to realistic values during calibration as was undertaken for the SM model. Therefore, they should probably be treated as “effective” parameters in the calibration process. Finally, underestimation of simplification error could also stem from interaction with boundary conditions.

The second type of ROM that was used in the study were ANNs. As a most fundamental difference, a single ANN model was created and calibrated for each single location in the data set. It was demonstrated that most predictions by ANN ROMs are generally free of bias and have small simplification errors associated to them. As all ANNs are calibrated against the respective time series from which those predictions come, this mainly proves that the chosen type of ANN is capable of reproducing such timelines adequately. From this, we conclude that where data is available, such ANNs can be used for short-term prediction and management purposes. The paired model analysis can aid this by providing bias correction, where necessary.

A direct comparison of the results for the two different ROMs is not expedient due to the different setup, prediction space and usage of calibration data sets. We conclude that the ANNs are formidable tools where

they are applicable: predictions for available data time series can be made with low simplification error and almost in real-time. For predictions not associated directly with measurements, a simple physically-distributed model like SM is the only choice, but the performance depends on how similar the prediction type is to the calibration data set. Some predictions of very different type or predictions dependent on parameter detail not represented in the simple model might be better assessed only relying on the complex model.Here, error can actually increase through the calibration of the simple model in comparison to complex model precalibration uncertainty estimates.

Overall, the presented methodology allows the evaluation of different types of ROMs in conjunction with the complex model and their classification into different areas of operation. The analysis of the strength and shortcomings of the simple models in regard to different predictions is adamant for their further potential use in management, be it scenario calculations or data worth analysis.

Acknowledgments

This work was funded by the German Research Foundation (grant WO 1781/1-1). The author would like to thank the Marlborough District Council for providing the data for this study and Mr. Saniruzzaman Suvo for his contribution to the simplification of the MODFLOW model.

Appendix A. Semi-random parameter field generation

To create (semi-)random pilot point parameters of the three aquifer layers (K_H and S_y fields), we considered two separate constraints that were informed by expert knowledge:

C1: Parameter ranges [p_{min}, p_{max}] as derived from local pumping tests, and

C2: Spatial correlation of parameters as defined by an exponential variogram (see Wöhling et al., 2018). The heterogeneity of a set of pilot point values that satisfy the variogram can be measured by a regularization function (e.g. as implemented by Doherty (2016) in PEST):

$$r = \sum \frac{1}{w_i}(p_j - p_k), \tag{22}$$

where p_j and p_k are any two pilot points within the range of the variogram, w_i is the weight of the ith combination of those two pilot points (which is proportional to their distance) and r is the regularization function value. Different random fields adhering to the variogram do not result in the same regularization function value r. We generate a probability density function (pdf) $f(\mathbf{r}_{vario})$ of possible variogram-adhering regularization function values $\mathbf{r}_{vario}$. Constraint C2 is defined as:

$$f(\mathbf{r}_{chosen}) \approx f(\mathbf{r}_{vario}), \tag{23}$$

where $f(\mathbf{r}_{chosen})$ is the pdf of the chosen parameter's regularization function values $\mathbf{r}_{chosen}$.

The following workflow was implemented to create 1000 random parameter sets for each of the parameter six fields in the complex model (3 *HC* layers + 3 *SY* layers) that satisfy both constraints C1 and C2:

1. Sampling of 1,000,000 random parameter sets $\mathbf{K}_{uni}$ from uniform distributions of [p_{min}, p_{max}] (Table 2) and calculation of their regularization function values $\mathbf{r}_{\mathbf{uni}}$. Note that these parameter sets $\mathbf{K}_{uni}$ satisfy constraint C1, but their regularization function pdf $f(\mathbf{r}_{uni}) \neq f(\mathbf{r}_{vario})$, thus not satisfying C2.
2. Calculation of the variogram regularization function pdf $f(\mathbf{r}_{vario})$. This is done by:
   (a) Generation of 10,000 random parameter fields $\mathbf{F}_{vario}$ satisfying the variogram via the *fieldgen* utility of PEST (Doherty, 2015).
   (b) Calculation of the regularization function values for the above 10,000 random fields, $\mathbf{r}_{vario}$, and their corresponding probability density function $f(\mathbf{r}_{vario})$.
3. Resampling of the 1,000,000 random parameter sets $\mathbf{K}_{uni}$ with $f(\mathbf{r}_{uni})$ from step 1 via importance resampling (cf. Gelman et al., 2004): 1000 (unique) final random parameter sets $\mathbf{K}_{final}$ are drawn from $\mathbf{K}_{uni}$ with weights of each of the $i = 1, \ldots, 1{,}000{,}000$ samples of $\mathbf{r}_{\mathbf{uni}}$ calculated as:

$$w^i(r^i_{uni}) = \frac{f(r^i_{vario})}{f(r^i_{uni})}. \tag{24}$$

The importance resampling results in $\mathbf{K}_{final} \subset \mathbf{K}_{uni}$ and $f(\mathbf{r}_{final}) \approx f(\mathbf{r}_{vario})$. The final parameter set $\mathbf{K}_{final}$ satisfies both constraints C1 and C2.

References

Aster, R., Borchers, B., Thurber, C., 2011. Parameter Estimation and Inverse Problems. Elsevier, Amsterdam, Netherlands.

Boyce, S.E., Nishikawa, T., Yeh, W.W.-G., 2015. Reduced order modeling of the Newton formulation of MODFLOW to solve unconfined groundwater flow. Adv. Water Resour. 83, 250–262. https://doi.org/10.1016/j.advwatres.2015.06.005.

Dagan, G., Fiori, A., Jankovic, I., 2013. Upscaling of flow in heterogeneous porous formations: critical examination and issues of principle. Adv. Water Resour. 51, 67–85. https://doi.org/10.1016/j.advwatres.2011.12.017.

Doherty, J., 2015. Groundwater Data Utilities. Watermark Numerical Computing.

Doherty, J., 2016. PEST: Model-Independent Parameter Estimation—User Manual, sixth ed. Watermark Numerical Computing.

Doherty, J., Christensen, S., 2011. Use of paired simple and complex models to reduce predictive bias and quantify uncertainty. Water Resour. Res. 47 (12), W12534. https://doi.org/10.1029/2011WR010763.

Fienen, M.N., Masterson, J.P., Plant, N.G., Gutierrez, B.T., Thieler, E.R., 2013. Bridging groundwater models and decision support with a Bayesian network. Water Resour. Res. 49 (10), 6459–6473. https://doi.org/10.1002/wrcr.20496.

Gelman, A., Carlin, J., Stern, H., Rubin, D., 2004. Bayesian Data Analysis, second ed. Chapman & Hall/CRC.

Gosses, M., Nowak, W., Wöhling, T., 2018. Explicit treatment for Dirichlet, Neumann and Cauchy boundary conditions in POD-based reduction of groundwater models. Adv. Water Resour. 115, 160–171. https://doi.org/10.1016/j.advwatres.2018.03.011.

von Gunten, D., Wöhling, T., Haslauer, C., Merchán, D., Causapé, J., Cirpka, O., 2014. Efficient calibration of a distributed PDE-based hydrological model using grid coarsening. J. Hydrol. 519, 3290–3304. https://doi.org/10.1016/j.jhydrol.2014.10.025.

Höge, M., Wöhling, T., Nowak, W., 2018. A primer for model selection: the decisive role of model complexity. Water Resour. Res. 54 (3), 1688–1715. https://doi.org/10.1002/2017WR021902.

Lødøen, O.P., Tjelmeland, H., 2010. Bayesian calibration of hydrocarbon reservoir models using an approximate reservoir simulator in the prior specification. Stat. Modell. 10 (1), 89–111. https://doi.org/10.1177/1471082X0801000106.

Mandel, J., 1982. Use of the singular value decomposition in regression analysis. Am. Stat. 36 (1), 15–24. https://doi.org/10.1080/00031305.1982.10482771.

Niswonger, R., Panday, S., Ibaraki, M., 2011. MODFLOW-NWT, a Newton Formulation for MODFLOW-2005. Technical Report 6-A37. U.S. Geological Survey Techniques and Methods.

Niswonger, R., Prudic, D., 2005. Documentation of the Stream Flow-Routing (SFR2) Package to Include Unsaturated Flow Beneath Streams - A Modification to SFR1. Technical Report 6-A13. U.S. Geological Survey Techniques and Methods.

Oladyshkin, S., Nowak, W., 2012. Data-driven uncertainty quantification using the arbitrary polynomial chaos expansion. Reliab. Eng. Syst. Saf. 106, 179–190. https://doi.org/10.1016/j.ress.2012.05.002.

Scheidt, C., Caers, J., Chen, Y., Durlofsky, L.J., 2011. A multi-resolution workflow to generate high-resolution models constrained to dynamic data. Comput. Geosci. 15 (3), 545–563. https://doi.org/10.1007/s10596-011-9223-9.

Schöniger, A., Wöhling, T., Samaniego, L., Nowak, W., 2014. Model selection on solid ground: rigorous comparison of nine ways to evaluate Bayesian model evidence. Water Resour. Res. 50 (12), 9484–9513. https://doi.org/10.1002/2014WR016062.

Siade, A.J., Putti, M., Yeh, W.W.-G., 2012. Reduced order parameter estimation using quasilinearization and quadratic programming. Water Resour. Res. 48 (6), W06502. https://doi.org/10.1029/2011WR011471.

Stanko, Z.P., Boyce, S.E., Yeh, W.W.-G., 2016. Nonlinear model reduction of unconfined groundwater flow using POD and DEIM. Adv. Water Resour. 97, 130–143. https://doi.org/10.1016/j.advwatres.2016.09.005.

Tait, A., Henderson, R., Turner, R., Zheng, X., 2006. Thin plate smoothing spline interpolation of daily rainfall for New Zealand using a climatological rainfall surface. Int. J. Climatol. 26 (14), 2097–2115. https://doi.org/10.1002/joc.1350.

Taormina, R., Wing Chau, K., Sethi, R., 2012. Artificial neural network simulation of hourly groundwater levels in a coastal aquifer system of the Venice Lagoon. Eng. Appl. Artif. Intell. 25 (8), 1670–1676. https://doi.org/10.1016/j.engappai.2012.02.009.

Ushijima, T.T., Yeh, W.W.-G., 2015. Experimental design for estimating unknown hydraulic conductivity in an aquifer using a genetic algorithm and reduced order model. Adv. Water Resour. 86, 193–208. https://doi.org/10.1016/j.advwatres.2015.09.029.

Watson, T.A., Doherty, J.E., Christensen, S., 2013. Parameter and predictive outcomes of model simplification. Water Resour. Res. 49 (7), 3952–3977. https://doi.org/10.1002/wrcr.20145.

Wöhling, T., Gosses, M.J., Wilson, S.R., Davidson, P., 2018. Quantifying river-groundwater interactions of New Zealand's gravel-bed rivers: the Wairau plain. Groundwater 56, 647–666. https://doi.org/10.1111/gwat.12625.

Wood, B., 2009. The role of scaling laws in upscaling. Adv. Water Resour. 32, 723–736. https://doi.org/10.1016/j.advwatres.2008.08.015.